
Research Article: New Research | Cognition and Behavior

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

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

Cite as: eNeuro 2021; 10.1523/ENEURO.0362-20.2021

Received: 19 August 2020

Revised: 14 April 2021

Accepted: 15 April 2021

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.eneuro.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Copyright © 2021 Reddy et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

1 **1. Manuscript Title:**

2 Representational content of oscillatory brain activity during object recognition: contrasting
3 cortical and deep neural network hierarchies

4

5 **2. Abbreviated Title:** Oscillatory brain signals in object recognition

6

7 **3. List all Author Names and Affiliations in order as they would appear in the published
8 article**

9

10 Leila Reddy^{1,2,*}, Radoslaw Martin Cichy³, Rufin VanRullen^{1,2}.

11 1 ANITI, Université de Toulouse 3 ,31052 Toulouse, France.

12 2 CNRS, Centre de Recherche Cerveau et Cognition (CerCo), 31052 Toulouse France.

13 3 Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany.

14 * Corresponding Author.

15

16 **4. Author Contributions:**

17 Conceptualization: LR and RV; Data Recording (MEG, fMRI): RMC; Data Analysis: LR; Writing—
18 Original Draft: LR; Writing—Review and Editing: LR, RMC, RV; Funding acquisition: LR, RMC, RV.

19 **5. Correspondence should be addressed to (include email address):**

20 Leila Reddy : leila.reddy@cns.fr

21

22 **6. Number of Figures:** 8

23 **7. Number of Tables:** 0

24 **8. Number of Multimedia:** 0

25 **9. Number of words for Abstract:** 218

26 **10. Number of words for Significance Statement:** 108

27 **11. Number of words for Introduction:** 885

28 **12. Number of words for Discussion:** 1356

29 **13. Acknowledgements:** n/a

30 **14. Conflict of Interest:** The authors have no competing interests.

31 **15. Funding sources:** Funded by an ERC Consolidator Grant P-CYCLES number 614244, OSCI-

32 DEEP ANR grant ANR-19-NEUC-0004, and ANITI ANR-19-PI3A-0004 to R.V., a NVIDIA GPU grant,

33 Vis-Ex ANR-12-JSH2-0004, and AI-REPS ANR-18-CE37-0007-01 to L.R, and an Emmy Noether grant

34 CI241/1-1 to RMC.

35

36

37 **Representational content of oscillatory brain activity during object recognition:**
38 **contrasting cortical and deep neural network hierarchies**

39 **Abstract:**

40 Numerous theories propose a key role for brain oscillations in visual perception. Most of these theories
41 postulate that sensory information is encoded in specific oscillatory components (e.g., power or phase)
42 of specific frequency bands. These theories are often tested with whole-brain recording methods of
43 low spatial resolution (EEG or MEG), or depth recordings that provide a local, incomplete view of the
44 brain. Opportunities to bridge the gap between local neural populations and whole-brain signals are
45 rare. Here, using representational similarity analysis in human participants we explore which MEG
46 oscillatory components (power and phase, across various frequency bands) correspond to low or high-
47 level visual object representations, using brain representations from fMRI, or layer-wise
48 representations in seven recent Deep Neural Networks (DNNs) as a template for low/high-level object
49 representations. The results showed that around stimulus onset and offset, most transient oscillatory
50 signals correlated with low-level brain patterns (V1). During stimulus presentation, sustained beta
51 (~20Hz) and gamma (>60Hz) power best correlated with V1, while oscillatory phase components
52 correlated with IT representations. Surprisingly, this pattern of results did not always correspond to
53 low- or high-level DNN layer activity. In particular, sustained beta-band oscillatory power reflected
54 high-level DNN layers, suggestive of a feed-back component. These results begin to bridge the gap
55 between whole-brain oscillatory signals and object representations supported by local neuronal
56 activations.

57 **Significance Statement:**

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

66 **Introduction:**

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

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

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

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

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

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

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

131 **Methods:**

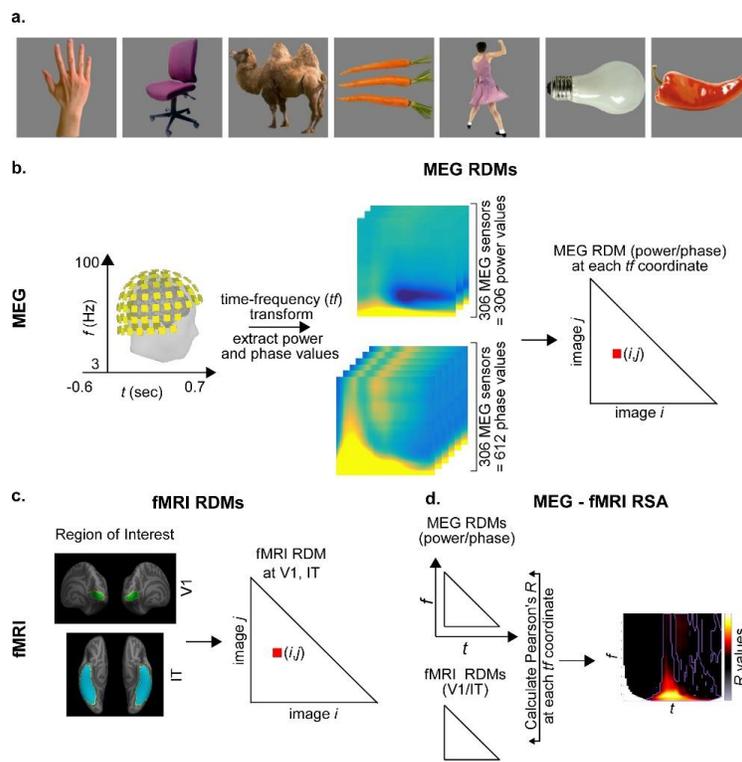
132 **Experimental paradigm and data acquisition:** The data analyzed in this study was obtained from
 133 (Cichy, Pantazis et al. 2014), and detailed methods can be obtained from that paper.

134 Fifteen human subjects of either sex performed separate MEG and fMRI sessions while they viewed a
 135 set of 92 images. The image set consisted of human and non-human faces and bodies, and artificial
 136 and natural everyday objects. The 92-image stimulus set was taken from the Kiani image set (Kiani,
 137 Esteky et al. 2007), which consists of cutout objects on a gray background (see examples in Figure
 138 1a).

139 In the MEG sessions, each image was presented for 0.5s followed by an inter-stimulus interval (ISI) of
 140 1.2 or 1.5s. Every 3-5 trials, a target paperclip object was presented, and subjects' task was to press a
 141 button and blink whenever they detected this target image. Subjects performed 2 MEG sessions, of 2
 142 hours each. In each session they performed between 10 to 15 runs. Each image was presented twice
 143 in each run, in random order.

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

148 The MEG data was acquired from 306 channels (204 planar gradiometers, 102 magnetometers,
 149 Elekta Neuromag TRIUX, Elekta, Stockholm) at the Massachusetts Institute of Technology. The MRI
 150 experiment was conducted on a 3T Trio scanner (Siemens, Erlangen, Germany), with a 32-channel
 151 head coil. The structural images were acquired using a T1-weighted sequence (192 sagittal slices,
 152 FOV = 256mm², TR=1,900ms, TE=2.52ms, flip angle=9 degrees). For the fMRI runs, 192 images were
 153 acquired for each participant (gradient-echo EPI sequence: TR = 2,000ms, TE=32 ms, flip angle = 80
 154 degrees, FOV read = 192 mm, FOV phase = 100%, ascending acquisition gap = 10%, resolution =
 155 2mm, slices=25).



156

157 **Figure 1:** MEG-fMRI RSA analysis. (a) examples from our 92-image set (b) MEG analysis and MEG
158 representational dissimilarity matrices (RDMs). From the MEG signals, the complex time frequency (TF) transform
159 was computed for each of the 306 MEG sensors. The amplitude and phase (separated into cosine and sine)
160 values were extracted from the complex number at each TF coordinate, and a MEG RDM was constructed,
161 reflecting the distance between oscillatory activation patterns for every pair of images (i,j) (see methods for
162 details). As a result, we obtained a power and phase MEG RDM at each TF coordinate for each participant. (c)
163 fMRI RDMs were obtained from (Cichy, Pantazis et al. 2014). Two regions of interest (ROI) were defined: V1 and
164 IT and one fMRI RDM was obtained for each ROI, and each participant, reflecting the distance between BOLD
165 activation patterns for every pair of images (i,j). (d) Representational Similarity Analysis (RSA) consists in
166 comparing two (or more) RDMs. The MEG power or phase RDMs were compared to the fMRI RDMs (V1 or IT) by
167 computing the partial Pearson's R . This step was performed at each TF coordinate, resulting in an RSA map of R
168 values at each TF coordinate, for each subject and ROI.

169
170 **MEG analysis - preprocessing:** MEG trials were extracted with a 600 ms baseline before stimulus
171 onset until 1200 ms post-stimulus onset. A total of 20-30 trials were obtained for each stimulus
172 condition, session, and participant. Each image was considered as a different stimulus condition.

173 Data were analyzed using custom scripts in Matlab (Mathworks) and FieldTrip (Oostenveld, Fries et al.
174 2011). Data were downsampled offline to 500 Hz. For each trial and sensor, we computed the
175 complex time frequency decomposition using multitapers. Parameters used were: 50 distinct
176 frequencies increasing logarithmically from 3 to 100 Hz, over a time interval of -600ms to 700ms with
177 respect to stimulus onset, in steps of 20 ms. The length of the sliding time window was chosen such
178 that there were two full cycles per time-window. The amount of smoothing increased with frequency
179 ($0.4 * \text{frequency}$).

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

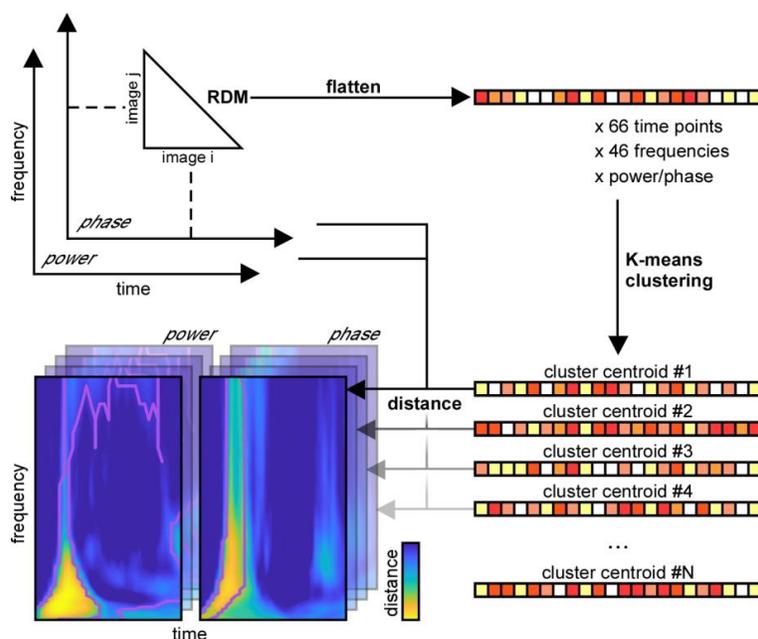
188 **MEG analysis – multivariate analysis (Figure 1b):** At each TF coordinate and for each stimulus
189 condition, we next arranged the 306 power values from the 306 MEG sensors into a 306-dimensional
190 vector representing the power pattern vector for that stimulus condition. Similarly, at each TF
191 coordinate and for each stimulus condition we concatenated the 306 sine and 306 cosine values into a
192 612-dimensional phase pattern vector for that stimulus condition.

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

199 **fMRI analysis (Figure 1c):** The preprocessing steps for the fMRI data are described in detail in
200 (Cichy, Pantazis et al. 2014). For the multivariate analysis, two regions of interest (ROIs) were defined:
201 V1 and IT. In each subject, for each ROI, voxel activation values were extracted for each stimulus
202 condition, and the resulting values were arranged in a pattern vector for each stimulus condition. Then,
203 in each ROI, for each pair of stimulus conditions, the corresponding pattern vectors were correlated
204 using the Pearson correlation measure, and the resulting 1-correlation value was assigned to the
205 92x92 fMRI RDM. For further analysis the fMRI RDMs were averaged across the 15 subjects, resulting
206 in one RDM per ROI. The fMRI RDMs were provided by R. Cichy, D. Pantazis and A. Oliva (Cichy,
207 Pantazis et al. 2014).

208 **MEG-fMRI Representational Similarity Analysis (RSA) (Figure 1d):** Representational Similarity
 209 Analysis (RSA) consists in comparing two (or more) RDMs. RSA between the MEG and fMRI RDMs
 210 was performed by computing the partial Pearson's correlation between each MEG (phase or power)
 211 RDM with each fMRI RDM (V1 or IT), while partialling out any contribution from the other fMRI RDM
 212 (IT or V1). We chose to perform a partial correlation because the V1 and IT RDMs were positively
 213 correlated with each other ($r \sim 0.3$); compared to a standard correlation, the partial correlation allowed
 214 us to isolate the unique correlation of each fMRI RDM with the MEG RDM, while discarding their joint
 215 contribution.

216 This procedure resulted in four RSA maps per subject (power/phase MEG RDMs x V1/IT fMRI RDMs).
 217 Each RSA map shows the R-value between the MEG signals and the V1/IT activation patterns at each
 218 TF point. Significance of the RSA result was evaluated with a paired t-test against 0, FDR corrected,
 219 $\alpha = 0.05$.



220

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

227 **Clustering analysis:** The MEG time-frequency RDMs are heavily correlated with each other. To
 228 facilitate the interpretation of the information content of oscillatory signals, and to determine which
 229 features co-vary and which are independent, K-means clustering was performed on the MEG power
 230 and phase RDMs. Clustering was performed on the 4186-dimensional $((92 \times 92 - 92)/2)$ RDMs across all
 231 (66) time points and (46) frequency points, combining the power and phase signals (resulting in
 232 $46 \times 66 \times 2 = 6072$ data points to cluster in a 4186-dimensional space). K-means was implemented with
 233 the Matlab function `kmeans`, with the correlational distance measure, five replicates, and the number
 234 of clusters going from 1 to 20. The optimal number of clusters was then determined with the elbow
 235 criterion defined as the point just before the local maximum of the second derivative of the residual
 236 sum of squares (corresponding to the point at which adding another cluster would only provide a
 237 marginal gain in variance explained). With this method, the first elbow occurred at $k=7$ clusters.

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

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

$$247 \quad L = \sigma \left(\frac{(R_{IT} - R_{V1})}{(R_{IT} + R_{V1})} \right) \quad (\text{eq. 1})$$

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

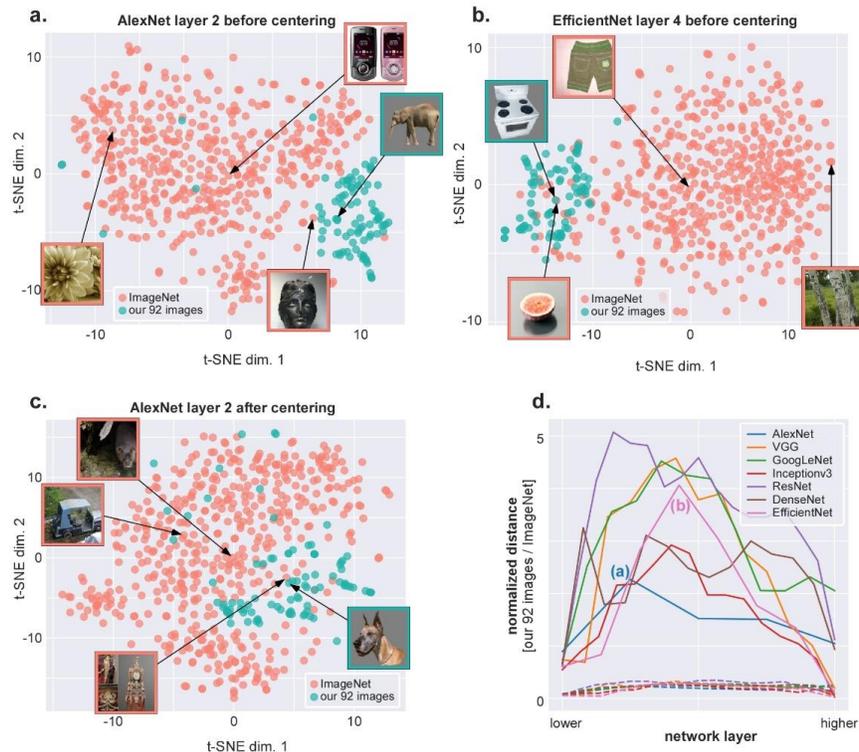
250 Significance of RSA between the cluster centroids and the fMRI RDMs was computed with a surrogate
 251 test. On each iteration, the cluster centroid RDM was randomly shuffled and the partial correlation was
 252 computed between this shuffled RDM and the true RDM. This procedure was repeated for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

255 **Deep Neural Network (DNN) RDMs:** The MEG phase/power representations were also compared to
 256 representations in seven DNNs (so as to ensure that conclusions were not dependent on one specific
 257 network architecture): AlexNet (Krizhevsky, Sutskever et al. 2012), VGG16 (Simonyan and Zisserman
 258 2014), GoogleNet (Szegedy, Liu et al. 2015), InceptionV3 (Szegedy, Ioffe et al. 2017), ResNet50 (He,
 259 Xiangyu et al. 2016), DenseNet121 (Huang, Liu et al. 2017) and EfficientNetB3 (Tan and Le 2019),
 260 processing the same 92 images as in our MEG and fMRI data. However, in contrast to our 92-image
 261 stimulus set, which consisted of cutout objects on a gray background, the DNNs had been trained on
 262 images from ImageNet (millions of photographs with one or more objects in natural backgrounds). The
 263 networks had thus learned optimal representations for their training set, but in this representation
 264 space our 92 images tended to cluster into a remote "corner" (Figure 3), with low dissimilarity (1-
 265 Pearson's R) values between images, and a resulting RDM of poor quality. To retrieve meaningful
 266 distances between the representations of the 92 images, we first performed a centering procedure: we
 267 centered the activation of each layer of each DNN by subtracting the mean activation of an
 268 independent set of 368 images from the Kiani image set. This independent image set consisted of four
 269 images from each of the categories in our 92-image set. Importantly, because the image set used for
 270 centering did not include any of the 92 images from our study, there was no circularity in the centering
 271 operation, nor any leakage of information between the representations of our 92 images. This
 272 centering procedure contributed to minimize the potential problems arising from differences between
 273 our dataset and the standard ImageNet dataset—but it did not completely alleviate these differences,
 274 as can be seen in Figure 3c.

275 RDMs were constructed for several convolutional layers of each network based on the layer activation
 276 values. There were 5 layers for AlexNet, 13 for VGG16, 12 for GoogleNet, 16 for InceptionV3, 17 for
 277 ResNet50 (hereafter referred to as ResNet), 14 for DenseNet121 (hereafter DenseNet) and 8 for
 278 EfficientNetB3 (hereafter EfficientNet). These layers were chosen so as to span the entire network
 279 hierarchy, without making the analysis computationally intractable (as some networks can contain
 280 more than 200 layers to choose from). RSA was then performed (with the Spearman correlation)
 281 between these RDMs and the centroid of each cluster (see above for details of the clustering
 282 analysis). The layer with maximum RSA, normalized by the number of layers for this DNN, was taken

283 to reflect the information content of this cluster between 0 (in the terminology defined in the
 284 Introduction, “low-complexity” corresponding to the DNN’s first layer) and 1 (“high-complexity”,
 285 corresponding to the DNN’s last layer), and finally averaged across the seven DNNs.

286 **Code Accessibility:** Custom code can be made available upon request.



287

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

310

311

312

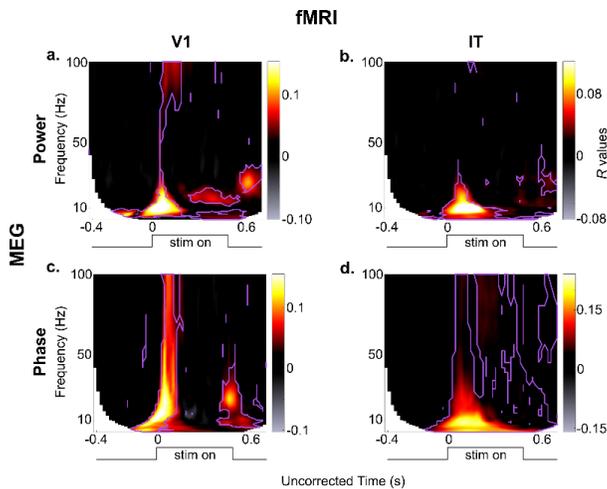
313 **Results:**

314 Fifteen participants viewed the same set of 92 images while fMRI and MEG data was recorded (in
315 separate sessions). The image set consisted of human and non-human bodies and faces, and artificial
316 and natural stimuli. Each stimulus was presented for 0.5s, followed by a 1.2 or 1.5s baseline period.

317 To assess oscillatory components, we extracted stimulus-related activity from -600ms to 1200ms
318 relative to stimulus onset from the MEG data. For each trial, and each sensor a time-frequency (TF)
319 decomposition was performed, and a power and phase value extracted at each time and frequency
320 point. These values were used to compute representational dissimilarity matrices (RDMs) at each TF
321 point, separately for power and phase (Methods and Figure 1a). Each element in the MEG RDMs
322 indicates how distinct the corresponding images are in the MEG power or phase spaces, and the
323 entire MEG RDM is a summary of how the 92-image stimulus set is represented in the MEG oscillatory
324 power or phase at each TF point.

325 To assess local patterns of neural activity we generated fMRI RDMs by performing comparisons
326 between the local BOLD activation patterns of pairs of images in V1 and IT (Cichy, Pantazis et al.
327 2014). Two fMRI RDMs were obtained (Figure 1b), one for V1 and one for IT. The fMRI RDMs are a
328 measure of the representation of the image set in the voxel space of V1 and IT local neural activity.

329



330

331

332 **Figure 4:** Results of the 2x2 RSA comparisons (MEG power/phase x fMRI V1/IT), averaged over all subjects. The
333 purple contours mark those regions in the maps that are significantly different from zero (paired *t*-test against 0
334 across $N=15$ subjects, FDR correction, $\alpha = 0.05$). Note that the absolute latencies are not directly comparable
335 across frequencies, because of different smoothing windows applied at the different frequencies when performing
336 the TF transform (hence, the x-axis is labeled as uncorrected time).
337

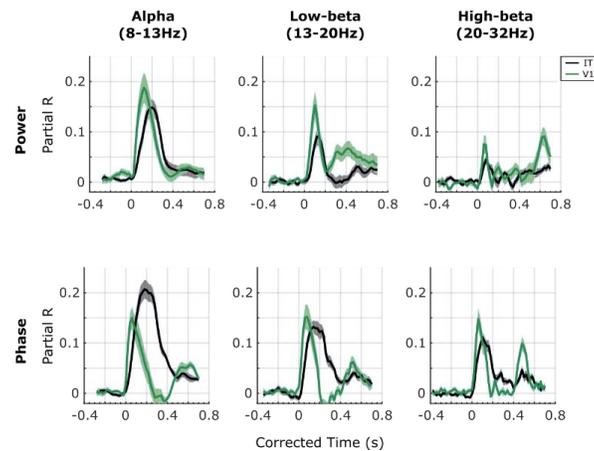
338 **Bridging the space, time and frequency gap in object recognition:**

339 How similar is the oscillatory representation of the images to their representation in each brain region?

340 The MEG RDMs (power and/or phase) at each TF point represent the stimulus set in a large-scale

341 brain oscillatory activity space, while the fMRI RDMs represent the same image set via BOLD activity
342 in a local population of neurons in two brain regions (V1 or IT). We evaluated the similarity of
343 representations in the time-frequency domain with those in the fMRI activation patterns by computing
344 the partial Pearson's correlation between the MEG RDMs (phase or power) with the fMRI RDMs
345 (V1/IT), at each TF point (Figure 1c). This analysis resulted in four time frequency maps of R-values
346 (or RSA maps), which provide the unique correspondence between whole-brain oscillations and local
347 patterns of neural activity in V1 and IT, at each TF point (Figure 4). With these maps we can ask if and
348 when stimulus information contained in oscillatory phase or power at each TF point resembles BOLD
349 activations in a given brain region (V1/IT), and potentially, which region it resembles more. The
350 advantage of measuring partial correlation (instead of a standard correlation) is to discard the
351 (potentially large) portion of the variance in oscillatory representations that is explained equally well by
352 V1 or IT BOLD representations—owing to the fact that V1 and IT signals already share similarities.
353 This way, we concentrate on the part of oscillatory representations that is *uniquely* explained by each
354 brain region-of-interest.

355 Our results show that different oscillatory components map to different brain regions at different
356 moments in time. Overall, the absolute maximum of representational similarity with brain area V1
357 occurred in the alpha band around 120ms post-stimulus for oscillatory power, whereas the absolute
358 maximum related to area IT occurred for theta- and alpha-phase around 200-300ms. More generally, a
359 strong increase in representational similarity was observed shortly after stimulus onset in all four
360 maps. The frequency, latency and duration of these similarity effects depended however on the exact
361 oscillatory signal (power, phase) and brain region (V1, IT). In terms of MEG power (Figure 4a, b), the
362 latencies (see also Figure 5) respected the hierarchical order of visual processing (Nowak and Bullier
363 1998) with an increase in representational similarity in the lower (<20 Hz) frequency bands occurring
364 around the evoked response first for RSA with V1, and about 20-30ms later for RSA with IT (paired t-
365 test against 0, FDR corrected, $\alpha=0.05$). This latency difference is similar to that reported in (Cichy,
366 Pantazis et al. 2014), where the peak correspondence between the average MEG signal and V1
367 activity occurs about 30 ms prior to the peak with IT activity. The onset response in the V1 RSA map
368 also consisted of high gamma frequencies (>70Hz), whereas this high-gamma activity was not
369 observed in the IT RSA map. This is compatible with recent findings showing that gamma oscillations
370 in early visual cortex are particularly prominent for certain stimuli, yet can be entirely absent for others
371 (Hermes, Miller et al. 2015). A sustained low-beta (20Hz, 200-500ms) and an offset high-beta (30Hz,
372 ~600ms) response also corresponded to V1 representations, although neither of these effects were
373 observed in the IT RSA map (see also Figure 5). In terms of stimulus representations in the MEG
374 oscillatory phase (Figure 4c, d), after an initial broadband (3-100Hz) transient peak at stimulus onset
375 corresponding to V1 representations, stimulus information carried by sustained oscillatory phase
376 resembled IT representations in the low (<20Hz) and high frequency (60Hz) bands, and this
377 resemblance persisted until the end of the trial. Phase representations corresponding to V1 patterns
378 were observed again around stimulus offset, at alpha (~10Hz) and beta (20-30Hz) frequencies (see
379 also Figure 5), in line with the known involvement of V1 neurons in OFF responses (Jansen, Jin et al.
380 2019).



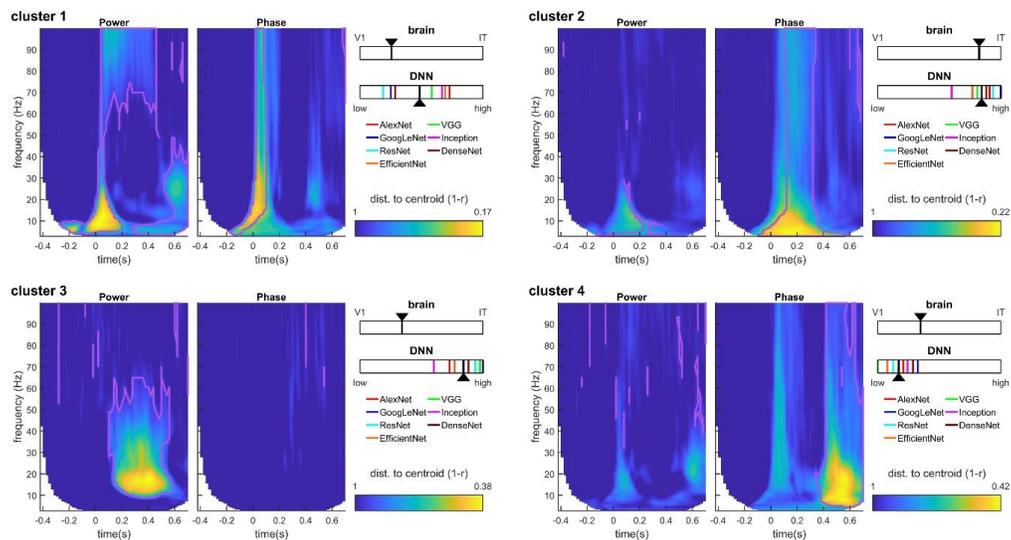
381

382 **Figure 5.** Profile of the results of the RSA with V1 (green lines) and IT (black lines) in oscillatory power (top row)
 383 and oscillatory phase (bottom row) in different frequency bands. To examine the RSA maps in more detail, we
 384 extracted their time courses in different traditional frequency bands: alpha (8-13 Hz), low-beta (13-20 Hz), and
 385 high-beta (20-32 Hz). In each of these frequency bands we computed the average R-values. Since the TF
 386 decomposition induces temporal smearing, and the amount of smearing differs for different frequencies, in order
 387 to interpret the latencies of the representational similarities, we corrected for this smearing effect. Specifically, to
 388 avoid underestimating the onset latencies, we corrected time by adding half the wavelet window duration at each
 389 frequency. Note that the same correction was applied to the two curves compared in each plot. Solid lines are the
 390 means across subjects, and the shaded areas correspond to the SEM across subjects.
 391

392 These results thus suggest that different oscillatory components correspond to different brain regions
 393 at different time-frequency points. However, since the RDMs in the time-frequency space are heavily
 394 correlated with each other, it is difficult to ascertain from this analysis which power/phase features co-
 395 vary, and which effects occur independently. To better interpret the results shown in Figure 4 we
 396 turned to a clustering analysis. The clustering analysis allowed us to reduce the dimensionality of the
 397 dataspace and to determine which oscillatory signals occurred jointly, and which are independent. We
 398 performed k-means clustering jointly on the power and phase RDMs. That is, each RDM (one for each
 399 time point, frequency, and phase/power signal), was considered as an input data point for the
 400 clustering analysis, which returned the corresponding cluster index assigned to each point (see Figure
 401 2). The results of the clustering analysis for $k=7$ clusters (the “optimal” number of clusters for our
 402 dataset) are shown in Figure 6 (see also Figure 7). The first cluster (ranked by smallest average
 403 distance from cluster centroid) corresponded to early broadband (0-100 Hz) phase and power RDMs,
 404 followed by sustained gamma power (>60Hz), and beta power (20-30Hz) at stimulus offset. The
 405 second cluster corresponded to broadband (0-100Hz) and sustained (0.1-0.4s) phase effects after
 406 stimulus onset, without any noticeable power effects. The third cluster consisted primarily of sustained
 407 (0.1-0.6s) beta (10-30Hz) and low-gamma (<60Hz) power, without any noticeable phase effects. The
 408 fourth cluster reflected broadband phase effects (0-100Hz) at stimulus offset (without associated
 409 power effects). The last 3 clusters (5-7) all displayed pre-stimulus effects in alpha-beta power, or alpha
 410 or gamma phase, characteristic of spontaneous, stimulus-unrelated activity that we did not investigate
 411 further (Figure 7). The clustering analysis performed on the MEG RDMs thus identified four main
 412 clusters of power and phase oscillatory components that occurred at different time points and in
 413 different frequency bands after stimulus onset.

414 How do the oscillatory representations in each cluster, and their different time and frequency profiles
 415 relate to local processing in V1 and IT as measured by fMRI representations? To address this
 416 question, we performed RSA between the cluster centroids and the V1 and IT RDMs. The cluster
 417 centroids correlated to different degrees with both V1 and IT (all partial R-values between 0.12 and
 418 0.49; all significant at $p < 1e-5$ with a surrogate test; see methods). To directly contrast the

419 representational similarity of each brain area (V1, IT) to the cluster centroid, we combined the two
 420 RSA partial R -values into a single scale (see Methods, equation 1). According to this scaling (see
 421 insets in Figure 6), the transient broadband phase and power effect with sustained gamma power in
 422 cluster 1 corresponded best with V1 representations (i.e., “low-level”). Conversely, the broadband
 423 sustained phase effects of cluster 2 corresponded best to IT representations (“high-level”). The other
 424 two clusters (sustained beta-gamma power in cluster 3, broadband offset-transient phase in cluster 4)
 425 had more balanced similarity to both V1 and IT, with a slight inclination towards V1. Thus, transient
 426 oscillatory components occurring around stimulus onset correspond more closely to V1
 427 representations, whereas the more sustained components could be either more IT-like, or less
 428 localized depending on the frequency of the oscillations. These results thus suggest a complex link
 429 between oscillatory representations and local processing in V1 or IT. To try to clarify these
 430 relationships we next turned to using deep neural networks as a template for object representations.



431

432 **Figure 6.** Clustering analysis. K-means clustering was performed on the MEG power and phase RDMs. Each
 433 time-frequency plot shows the distance of each RDM from the centroid of the corresponding cluster. The purple
 434 lines correspond to the cluster boundaries as returned by the k-means algorithm, indicating that all points within
 435 the purple lines are assigned to this specific cluster based on their distance to the different cluster centroids. The
 436 distance to centroid (color scale) reflects how “stereotypical” each RDM is for the corresponding cluster (i.e., how
 437 close to the cluster centroid), a continuous scale that complements the discrete cluster assignment. For example,
 438 even though cluster 3 simultaneously encompasses oscillatory power across many frequencies from 10Hz to
 439 65Hz, we can see that low-beta frequencies (13-20Hz) are the most “stereotypical” for this cluster. The insets
 440 show the relative degree of RSA between the cluster centroid and V1/IT (top), or the cluster centroid and the DNN
 441 layer hierarchy (bottom). For the DNNs, the layer with maximum RSA, normalized by the number of layers in the
 442 DNN hierarchy, and averaged across the seven DNN types (colored ticks), was taken as the layer that
 443 corresponded to each cluster centroid (black arrowhead).

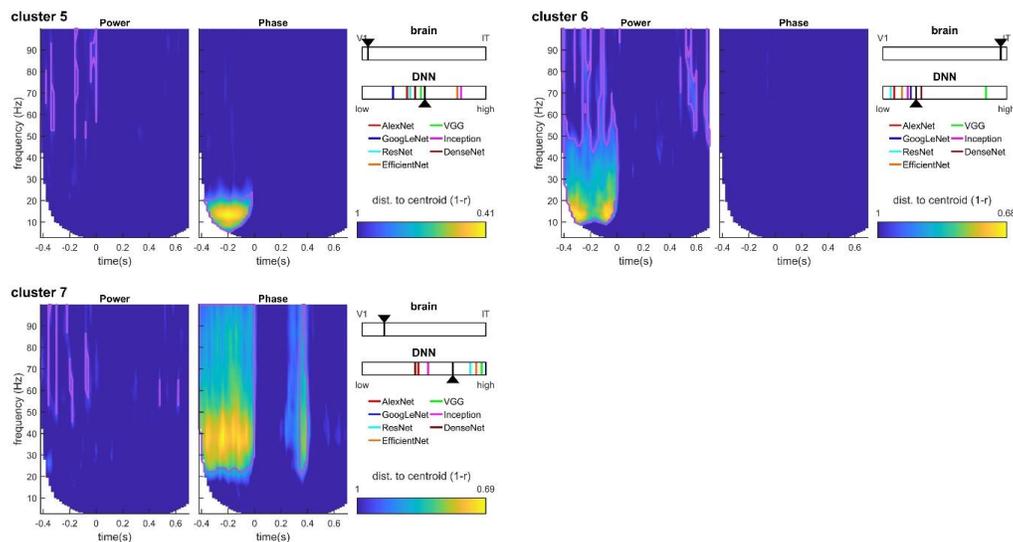
444

445 **Assessing representational complexity with deep neural networks:**

446 The fMRI RDMs are a representation of the image set in the multi-voxel space of V1 and IT. However,
 447 these fMRI representations are static because the fMRI BOLD signal used to construct the RDMs was
 448 measured over a period of several seconds. Neuronal activity in these regions, on the other hand, is
 449 known to evolve over fairly rapid timescales, on the order of hundreds of milliseconds as a result of
 450 feedback and top-down signals (Roelfsema, Lamme et al. 1998, Lamme and Roelfsema 2000). The
 451 fMRI RDMs are thus limited representations of the image set, potentially mixing low- and high-
 452 complexity brain activity from different moments in each trial. Therefore, while it is tempting to interpret

453 the oscillatory signals composing cluster 1 as low-complexity, because they are more V1-like, and
 454 those forming cluster 2 as high-complexity (more IT-like), such a conclusion would be premature as it
 455 ignores the dynamics of neural responses within and across brain regions, and how these neural
 456 responses evolve over different timescales. To obtain a complementary picture of low and high-level
 457 object representations, we considered the representations of our image set in different layers of feed-
 458 forward deep neural networks (DNNs) pretrained on a large dataset of natural images. To ensure the
 459 generality of our results we assessed seven different DNNs: AlexNet (Krizhevsky, Sutskever et al.
 460 2012), VGG16 (Simonyan and Zisserman 2014), GoogleNet (Szegedy, Liu et al. 2015), InceptionV3
 461 (Szegedy, Ioffe et al. 2017), ResNet (He, Xiangyu et al. 2016), DenseNet (Huang, Liu et al. 2017) and
 462 EfficientNet (Tan and Le 2019). Activity in each layer of these DNNs is not influenced by top-down or
 463 recurrent connections, and consequently represents a truly hierarchical evolution in the complexity of
 464 image representations, from low- to high-complexity. Indeed, several studies have suggested that
 465 DNN representations approximate the feed-forward cascade of the visual processing hierarchy in the
 466 brain (Khaligh-Razavi and Kriegeskorte 2014, Cichy, Khosla et al. 2016). Performing RSA between
 467 MEG oscillatory RDMs and DNN layer RDMs should thus reveal which features of the MEG oscillatory
 468 representations correspond to low- vs high-complexity object representations.

469 An RDM was obtained for several representative convolutional layers of the seven DNNs. RSA was
 470 then performed between the cluster centroids of the MEG RDMs and the DNN RDMs. For each cluster
 471 and DNN, the layer with maximum RSA was determined, and scaled between 0 (lowest layer, low-
 472 complexity information) and 1 (highest layer, high-complexity information) based on the number of
 473 layers in the DNN hierarchy. Despite notable differences between the seven DNNs, the analysis
 474 revealed that clusters 2 and 3 mapped best to higher DNN layers, cluster 1 to intermediate layers, and
 475 only cluster 4 had similarity to lower layers. This is in stark contrast with the results of fMRI RSA,
 476 which had ranked clusters 2, 4, 3 and 1 in order of decreasing complexity. The most striking difference
 477 is obtained for cluster 3 (sustained beta-gamma power): a high-complexity representation according to
 478 DNNs, but closer to V1 than to IT according to fMRI. Based on the logic above, this cluster is likely to
 479 reflect feed-back signals that carry high-complexity object information (visible in high DNN layers)
 480 down to lower brain regions (visible in V1 BOLD signals).



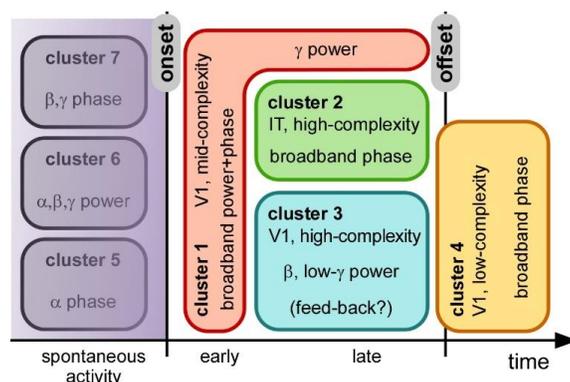
481

482 **Figure 7.** Clustering results for clusters 5-7. We identify these clusters as noise components because (i) their
 483 distance to the cluster centroid is typically higher than for other clusters, and (ii) they mainly map onto pre-
 484 stimulus oscillatory activity. Pre-stimulus oscillations, while accounting for a sizeable portion of the (notoriously
 485 noisy) MEG signal variance, cannot possibly encode the identity of a stimulus that has not been presented yet.

486 Pre-stimulus alpha is a well-studied oscillatory component reflecting the attention state of the observer, and
 487 whose phase is known to modulate the subsequent ERP amplitudes and latencies; as such, it is not surprising
 488 that the phase of this oscillatory component would induce a separate cluster of RDM patterns (cluster 5).
 489 Similarly, pre-stimulus alpha-beta power (cluster 6) and gamma phase (cluster 7) could reflect preparatory
 490 attention or motor signals (including muscular artifacts) not related to stimulus identity. Notations as in Figure 6.
 491 Note that for consistency with the previous figure, we continue to report the V1-IT RSA scaling value in the insets;
 492 however, the corresponding correlation values were systematically lower for these clusters, and should thus be
 493 interpreted with caution (V1 partial correlations for clusters 5-7: [0.04, 0.00, 0.08], IT partial correlations: [0.00,
 494 0.04, 0.03]). In comparison, V1 partial correlations for clusters 1-4 ranged from 0.14 to 0.43, and IT partial
 495 correlations from 0.12 to 0.44. Similarly, we also report the DNN layer for which the correlation to the cluster
 496 centroid was maximal; however, this maximal correlation was consistently lower than in the previous figure, as
 497 expected for a pre-stimulus component (peak correlations averaged across DNNs for clusters 5-7: [0.02, 0.03,
 498 0.16], compared to values ranging from 0.24 to 0.45 for clusters 1-4).
 499

500 Discussion:

501 Our results (summarized in Figure 8) show that MEG oscillatory components at different frequencies
 502 carry stimulus-related information at specific times, which can be linked, via RSA, to stimulus
 503 representations in different brain regions (V1, IT), and with different representational complexity (as
 504 measured by deep neural networks). Importantly, the representational dynamics of brain oscillations
 505 can be very differently expressed by power vs. phase signals. At stimulus onset and offset, broadband
 506 phase transients (possibly related to fluctuations in evoked potential latencies) carry mainly low-
 507 complexity or intermediate-complexity information (clusters 1, 4 in Figure 6). However, during stimulus
 508 presentation, sustained phase information is visible across all frequencies, and consistently maps to
 509 high-level and high-complexity representations (IT and high DNN layers, cluster 2). Oscillatory power
 510 components (clusters 1 and 3) tend to correlate with both V1 and IT fMRI representations (with an
 511 inclination towards V1); however, onset-transient low-frequency (<20Hz) power together with
 512 sustained high-frequency (>60Hz) power (i.e., cluster 1) correspond best to intermediate DNN layers,
 513 whereas sustained beta-gamma power (20-60Hz) clearly maps to the highest DNN layers (cluster 3).



514

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

518 It is important to note that some of the TF components revealed here could be specific to the
 519 conditions of our experiment. For example, brain oscillations are often modulated by the subjective
 520 state, the participant's attention or the task instructions. As such, it is likely that different oscillatory
 521 patterns would be obtained for tasks involving active behavior rather than passive viewing of the
 522 images. Similarly, spurious muscular activity (including ocular saccades or micro-saccades) could be
 523 an important contribution to the observed TF components (Yuval-Greenberg, Tomer et al. 2008), as
 524 long as this activity would systematically differ between the different stimulus classes.

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

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

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

556 Our results highlight the importance of complementing MEG-fMRI RSA with another measure of
557 representational content such as feed-forward DNNs (Cichy, Khosla et al. 2016, Bankson, Hebart et
558 al. 2018, Hebart, Bankson et al. 2018, Khaligh-Razavi, Cichy et al. 2018). fMRI BOLD signals are
559 often analyzed such that they reflect a single static representation. Thus, they cannot distinguish
560 dynamics in local patterns as for example early feedforward and later feedback activity. By design,
561 feedforward DNN layers cannot be dynamically influenced by feedback signals, and could be
562 considered to provide a template for low- vs. high-complexity representations during the different
563 stages of image processing. Perhaps the best illustration of this notion stems from the discrepancy
564 between fMRI and DNN RSA for MEG cluster 3, which suggests that sustained beta-gamma power
565 during stimulus presentation could reflect feedback signals: best corresponding to V1 fMRI activity
566 (low-level), but higher DNN layers (high-complexity). Without this additional information (e.g., looking
567 at Figure 4a alone), one might have interpreted sustained beta-power as a strictly low-level signal. The
568 observed distinction between sustained power effects at lower frequencies (beta and low-gamma,
569 cluster 3) vs. higher frequencies (high-gamma, cluster 1) is consistent with a large number of recent
570 studies that reported a functional distinction between gamma-band and beta-band signals,
571 respectively supporting feed-forward and feedback communication (Fontolan, Morillon et al. 2014, van
572 Kerkoerle, Self et al. 2014, Bastos, Vezoli et al. 2015, Michalareas, Vezoli et al. 2016). Future work
573 could attempt to separate feedforward from feedback signals (e.g. with backward masking and/or

574 layer-specific fMRI), to confirm the differential contribution of gamma and beta-band oscillatory
575 frequencies to feedforward vs. feedback object representations, as determined with RSA.

576 In addition to their involvement in the transmission of feedforward and feedback signals, several
577 studies have shown that different oscillatory signals can carry distinct information about stimulus
578 properties (Smith, Gosselin et al. 2006, Romei, Driver et al. 2011, Schyns, Thut et al. 2011, Mauro,
579 Raffone et al. 2015). Here we considered whether oscillatory components in different frequencies
580 correspond to lower- or higher-complexity stimulus processing stages. Our results suggest that most
581 oscillatory brain activity, at least at the broad spatial scale that is measured with MEG, reflects already
582 advanced stimulus processing in object detection tasks. This result can be seen in Figure 6 where
583 most oscillatory components are more related to higher-level DNN layer representations, with the
584 exception of the offset-transient (cluster 4). Indeed, one might have expected that stimulus
585 representations at both stimulus onset and offset are more reflective of transient low-level and low-
586 complexity processing. However, while both onset and offset signals (cluster 1 and cluster 4) are
587 better matched to V1 than IT ("low-level", see also Figure 4c, d), in terms of DNN activations the
588 offset-transient (cluster 4) appears to be of much lower-complexity and the onset-transient of higher-
589 complexity (cluster 1). A tentative explanation could be that the continued presence of the stimulus
590 after the onset-transient supports a rapid refinement of object representations, which would not be the
591 case for the offset-transient (because the stimulus is absent from the retina). Indeed, it is remarkable
592 that, aside from this offset-transient broadband phase activity (cluster 4), no other oscillatory signal
593 was found to reflect low-level DNN layers (i.e., low-complexity information).

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

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

603

604 **References:**

- 605 Bankson, B. B., M. N. Hebart, I. I. A. Groen and C. I. Baker (2018). "The temporal evolution of
606 conceptual object representations revealed through models of behavior, semantics and deep
607 neural networks." *Neuroimage* **178**: 172-182.
- 608 Bastos, A. M., J. Vezoli, C. A. Bosman, J. M. Schoffelen, R. Oostenveld, J. R. Dowdall, P. De Weerd, H.
609 Kennedy and P. Fries (2015). "Visual areas exert feedforward and feedback influences
610 through distinct frequency channels." *Neuron* **85**(2): 390-401.
- 611 Cichy, R. M., A. Khosla, D. Pantazis, A. Torralba and A. Oliva (2016). "Comparison of deep neural
612 networks to spatio-temporal cortical dynamics of human visual object recognition reveals
613 hierarchical correspondence." *Sci Rep* **6**: 27755.
- 614 Cichy, R. M., D. Pantazis and A. Oliva (2014). "Resolving human object recognition in space and
615 time." *Nat Neurosci* **17**(3): 455-462.
- 616 Fontolan, L., B. Morillon, C. Liegeois-Chauvel and A. L. Giraud (2014). "The contribution of
617 frequency-specific activity to hierarchical information processing in the human auditory
618 cortex." *Nat Commun* **5**: 4694.
- 619 Fries, P. (2005). "A mechanism for cognitive dynamics: neuronal communication through
620 neuronal coherence." *Trends Cogn Sci* **9**(10): 474-480.

- 621 He, K., Z. Xiangyu, R. Shaoqing and S. Jian (2016). Deep residual learning for image recognition.
622 Proceedings of the IEEE conference on computer vision and pattern recognition.
- 623 Hebart, M. N., B. B. Bankson, A. Harel, C. I. Baker and R. M. Cichy (2018). "The representational
624 dynamics of task and object processing in humans." Elife **7**.
- 625 Hermes, D., K. J. Miller, B. A. Wandell and J. Winawer (2015). "Stimulus Dependence of Gamma
626 Oscillations in Human Visual Cortex." Cereb Cortex **25**(9): 2951-2959.
- 627 Huang, G., Z. Liu, L. Van Der Maaten and K. Q. Weinberger (2017). Densely connected
628 convolutional networks. Proceedings of the IEEE conference on computer vision and pattern
629 recognition.
- 630 Jansen, M., J. Jin, X. Li, R. Lashgari, J. Kremkow, Y. Bereshpolova, H. A. Swadlow, Q. Zaidi and J. M.
631 Alonso (2019). "Cortical Balance Between ON and OFF Visual Responses Is Modulated by the
632 Spatial Properties of the Visual Stimulus." Cereb Cortex **29**(1): 336-355.
- 633 Jensen, O. and A. Mazaheri (2010). "Shaping functional architecture by oscillatory alpha activity:
634 gating by inhibition." Front Hum Neurosci **4**: 186.
- 635 Kar, K., J. Kubilius, K. Schmidt, E. B. Issa and J. J. DiCarlo (2019). "Evidence that recurrent circuits
636 are critical to the ventral stream's execution of core object recognition behavior." Nat
637 Neurosci **22**(6): 974-983.
- 638 Khaligh-Razavi, S.-M. and N. Kriegeskorte (2014). "Deep supervised, but not unsupervised,
639 models may explain IT cortical representation." PLoS Comput Biol **10**(11): e1003915.
- 640 Khaligh-Razavi, S. M., R. M. Cichy, D. Pantazis and A. Oliva (2018). "Tracking the Spatiotemporal
641 Neural Dynamics of Real-world Object Size and Animacy in the Human Brain." J Cogn
642 Neurosci **30**(11): 1559-1576.
- 643 Kiani, R., H. Esteky, K. Mirpour and K. Tanaka (2007). "Object category structure in response
644 patterns of neuronal population in monkey inferior temporal cortex." J Neurophysiol **97**(6):
645 4296-4309.
- 646 Kietzmann, T. C., C. J. Spoerer, L. K. A. Sorensen, R. M. Cichy, O. Hauk and N. Kriegeskorte (2019).
647 "Recurrence is required to capture the representational dynamics of the human visual
648 system." Proc Natl Acad Sci U S A **116**(43): 21854-21863.
- 649 Kriegeskorte, N., M. Mur and P. Bandettini (2008). "Representational similarity analysis -
650 connecting the branches of systems neuroscience." Front Syst Neurosci **2**: 4.
- 651 Krizhevsky, A., I. Sutskever and G. E. Hinton (2012). "Imagenet classification with deep
652 convolutional neural networks." Advances in Neural Information Processing Systems **25**:
653 1097-1105.
- 654 Kuzovkin, I., R. Vicente, M. Petton, J. P. Lachaux, M. Baciú, P. Kahane, S. Rheims, J. R. Vidal and J.
655 Aru (2018). "Activations of deep convolutional neural networks are aligned with gamma
656 band activity of human visual cortex." Commun Biol **1**: 107.
- 657 Lamme, V. A. and P. R. Roelfsema (2000). "The distinct modes of vision offered by feedforward
658 and recurrent processing." Trends Neurosci **23**(11): 571-579.
- 659 Mauro, F., A. Raffone and R. VanRullen (2015). "A bidirectional link between brain oscillations
660 and geometric patterns." J Neurosci **35**(20): 7921-7926.
- 661 Michalareas, G., J. Vezoli, S. van Pelt, J. M. Schoffelen, H. Kennedy and P. Fries (2016). "Alpha-Beta
662 and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual
663 Cortical Areas." Neuron **89**(2): 384-397.
- 664 Nowak, L. and J. Bullier (1998). The timing of information transfer in the visual system. Cerebral
665 Cortex. J. H. Kaas, K. Rockland and A. Peters. New York, Plenum Press: 205-241.
- 666 Oostenveld, R., P. Fries, E. Maris and J. M. Schoffelen (2011). "FieldTrip: Open source software for
667 advanced analysis of MEG, EEG, and invasive electrophysiological data." Comput Intell
668 Neurosci **2011**: 156869.
- 669 Roelfsema, P. R., V. A. Lamme and H. Spekreijse (1998). "Object-based attention in the primary
670 visual cortex of the macaque monkey." Nature **395**(6700): 376-381.
- 671 Romei, V., J. Driver, P. G. Schyns and G. Thut (2011). "Rhythmic TMS over parietal cortex links
672 distinct brain frequencies to global versus local visual processing." Curr Biol **21**(4): 334-337.
- 673 Schyns, P. G., G. Thut and J. Gross (2011). "Cracking the code of oscillatory activity." PLoS Biol
674 **9**(5): e1001064.

- 675 Simonyan, K. and A. Zisserman (2014). "Very Deep Convolutional Networks for Large-Scale
676 Visual Recognition." CoRR **abs/1409.1556**.
- 677 Smith, M. L., F. Gosselin and P. G. Schyns (2006). "Perceptual moments of conscious visual
678 experience inferred from oscillatory brain activity." Proc Natl Acad Sci U S A **103**(14): 5626-
679 5631.
- 680 Szegedy, C., S. Ioffe and V. Vanhoucke (2017). "Inception-v4, Inception-ResNet and the Impact of
681 Residual Connections on Learning." Thirty-first AAAI conference on artificial intelligence.
- 682 Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A.
683 Rabinovich (2015). "Going deeper with convolutions." Proceedings of the IEEE Conference
684 on Computer Vision and Pattern Recognition: 1-9.
- 685 Tan, M. and Q. Le (2019). Efficientnet: Rethinking model scaling for convolutional neural
686 networks. International Conference on Machine Learning PMLR.
- 687 van Kerkoerle, T., M. W. Self, B. Dagnino, M. A. Gariel-Mathis, J. Poort, C. van der Togt and P. R.
688 Roelfsema (2014). "Alpha and gamma oscillations characterize feedback and feedforward
689 processing in monkey visual cortex." Proc Natl Acad Sci U S A **111**(40): 14332-14341.
- 690 Yuval-Greenberg, S., O. Tomer, A. S. Keren, I. Nelken and L. Y. Deouell (2008). "Transient induced
691 gamma-band response in EEG as a manifestation of miniature saccades." Neuron **58**(3): 429-
692 441.
- 693