Using DNNs as a yardstick for estimating the representational value of oscillatory brain signals.

Leila Reddy (leila.reddy@cnrs.fr)
CerCo, CNRS, Université de Toulouse
Toulouse, 31055 (France)

Radoslaw Martin Cichy (rmcichy@zedat.fu-berlin.de)

Rufin VanRullen (rufin.vanrullen@cnrs.fr)

Abstract:

Cognitive neuroscience must evaluate the information content and representational complexity of each brain signal: does it covary with physical attributes of sensory inputs (e.g. contrast, orientation), or with more elaborate attributes such as an object's category? Comparing response patterns between brain regions and/or recording modalities (e.g. MEG, fMRI) is a useful approach, but somewhat limited by the complexity of brain dynamics (e.g. "low-level" brain regions initially respond according to physical attributes, but are later affected by object category). Here, we followed recent studies that used feed-forward deep neural networks (DNNs) as a yardstick for the representational content of brain signals. We analyzed MEG oscillations, recorded while human subjects viewed images from different object categories. The multivariate response pattern for phase and amplitude signals in each oscillatory band (theta, alpha, beta, gamma) was compared with each layer of two standard DNNs (GoogLeNet, VGG) presented with the same object images. Overall, these large-scale oscillatory brain signals tended to coincide better with higher DNN processing layers; this was most evident for phase compared to amplitude, and for lower frequencies (<13Hz, theta and alpha). In contrast, high-frequency (~40Hz, beta and gamma) amplitude was the only oscillatory signal that best matched lower DNN layers.

Keywords: oscillations; MEG; theta; phase; amplitude; DNNs; representational similarity analysis;

Introduction

Brain oscillations are thought to play an important functional role in sensory perception and cognition. However, in relating computational mechanisms to oscillatory brain signals (from intracranial electrophysiological recordings or from scalp-level sensors such as EEG or MEG), a clear limitation is that the time or place at which signals are recorded does not directly indicate the nature of information encoded. Certain brain regions encode both low-level properties (e.g. contrast, orientation) and higher-level attributes of visual inputs (e.g. object category) at different times, while certain high-level regions can display response latencies as short as those observed in low-level regions. One strategy to overcome these obstacles is to rely on deep neural networks (DNNs) as an "objective" hierarchy of visual processing levels (Guclu & van Gerven, 2015; Cichy et al., 2016). These networks can perform object recognition with human-level accuracy, and display a selectivity to features of increasing complexity (from oriented edges to real object classes) not unlike that observed in neurophysiological experiments. However, because of their feed-forward architecture, the hierarchical rank of each DNN layer can be directly and unambiguously mapped onto a corresponding rank in representational feature complexity. Here, we applied a similar comparison strategy to help characterize the functional role of brain oscillations in various standard frequency bands: theta (4-8Hz), alpha (8-13Hz), low beta (13-20Hz), high beta (20-32Hz), low gamma (32-50Hz) and high-gamma (50-100Hz). We separately considered the information conveyed by oscillatory phase and amplitude.

Methods

MEG: Fifteen subjects were tested with MEG while they viewed 92 different objects presented at the center of the screen for 0.5s (Cichy, Pantazis & Oliva, 2014). Wavelet time-frequency (TF) decomposition for each trial and MEG sensor was performed at frequencies between 3-100Hz and from -0.6 to +0.7s relative to stimulus onset. MEG representational dissimilarity matrices (RDMs; Kriegeskorte, Mur & Bandettini, 2008) were computed separately for oscillatory amplitude and phase at each TF coordinate. Amplitude RDMs reflect how distinct the amplitude distribution across MEG channels is for each pair of stimuli. Phase RDMs represent to what extent each image in a pair is associated with its own distinct phase pattern.

DNNs: The GoogLeNet (Szegedy et al, 2015) and VGG (Simonyan & Zisserman, 2014) networks (Tensorflow implementations for Python) were presented with the same 92 images as the human subjects, and activation maps for each layer were extracted in order to compute RDMs. To make the networks comparable, we discarded the fully connected layers from the VGG network, and limited our analysis to the first 12 convolutional layers of both networks.

Representational Similarity Analysis (RSA): Next we computed a representational similarity analysis (RSA; Kriegeskorte et al, 2008) between each subject's MEG RDMs and the RDM for each layer of the DNNs. This RSA analysis results in a correlation value (r) for each DNN layer at each TF point. For each subject and within each oscillatory frequency band we extracted the maximum r value over the stimulus presentation period. Finally, for each subject and frequency band, we computed the regression slope of the

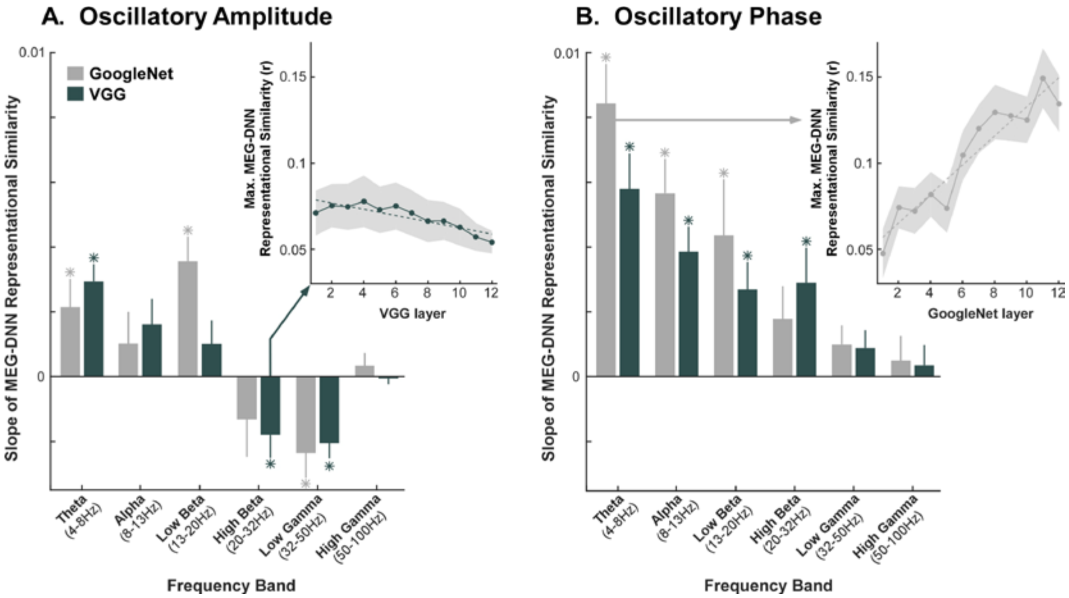

Figure 1. Quantifying representational content by comparing DNN layers and MEG oscillatory amplitude (A) or phase (B). Representational dissimilarity matrices (RDMs) were computed for each DNN layer and each oscillatory signal for each subject (N=15) at each time point during stimulus presentation. Correlating the MEG and DNN RDMs yielded a measure of representational similarity (r); we selected the maximum r value across time points. The insets show two examples of representational similarity data for high-beta amplitude/VGG (left), and for theta phase/GoogLeNet (right). The shaded regions represent s.e.m. across subjects. The slope of the representational similarity curves is the final measure used to quantify representational content (bar graphs: error bars represent s.e.m., * symbols denote a slope significantly different from zero, one-sample t-test, p<0.05). High slopes indicate a representation that preferentially encodes object features and category, while lower or even negative slopes suggest a representation that emphasizes physical attributes (e.g. contrast, orientation).

maximum r values across layers. That is, we do not consider absolute similarity between MEG data and each DNN layer, but rather how this similarity evolves across DNN layers.

Results and Conclusions

Results are summarized in Figure 1. Overall, the vast majority of MEG oscillatory brain signals displayed positive RSA slopes, i.e. their representational value matched the higher DNN layers better than the lower ones. The slopes were higher for oscillatory phase than amplitude, and were inversely related to oscillatory frequency. The highest representational value was found for the phase of theta-band oscillations (4-8Hz, see inset in Figure 1B). In contrast, only oscillatory amplitude signals in the high-beta and low-gamma bands (20-50Hz) showed negative RSA slopes (inset in Figure 1A), meaning that their representational content better matched the lower DNN layers.

MEG oscillations are global, complex time-varying signals that can be difficult to apprehend in a computational sense. Here, we showed that DNNs can serve as a yardstick to facilitate this endeavor. We highlighted beta-gamma amplitude and theta phase as two extremes of a continuum of representational value. Future work could examine the temporal dynamics of representational similarity between brain oscillations and DNNs.

Acknowledgments

This work was funded by an ERC Consolidator Grant P-CYCLES number 614244 to RV, and an Emmy Noether grant CI241/1-1 to RMC.

References

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. Nat Neurosci, 17(3), 455-462.

Cichy, R. M., 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.

Guclu, U., & van Gerven, M. A. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. J Neurosci, 35(27), 10005-10014.

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

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, abs/1409.1556.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-9).