Neural Dimensionality and Temporal Dynamics of Visual Representations

Zirui Chen (zchen160@jhu.edu)
Department of Cognitive Science, Johns Hopkins University
3400 N. Charles Street, Baltimore, MD 21218

Leyla Isik (lisik@jhu.edu)
Department of Cognitive Science, Johns Hopkins University
3400 N. Charles Street, Baltimore, MD 21218

Michael F. Bonner (mfbonner@jhu.edu)
Department of Cognitive Science, Johns Hopkins University
3400 N. Charles Street, Baltimore, MD 21218

Abstract

Understanding the temporal dynamics of visual representations in the brain is a fundamental challenge. While research has shown that neural time series data contain rich information where many visual features can be decoded, less is known about how the stimulus representation itself evolves over time and how these dynamics are related to feature decodability. Here, we investigated these questions using EEG recordings from subjects viewing everyday objects. We found that the dimensionality of stimulus-related variance rapidly increases to nearly full rank after stimulus onset and is sustained for several hundred milliseconds. During this time, the underlying representations oscillate, with every latent dimension undergoing multiple sign flips. Interestingly, the time course of feature decodability closely corresponds to the window of high-dimensionality, and temporal-generalization patterns of above- and below-chance decoding accuracies correspond to sign flips of the representational dimensions. Furthermore, we found that behavioral features and neural network representations each capture only a subset of the neural dimensionality, suggesting that significant portions of neural activity represent information not accounted for by current measures. Together, our findings show that natural images elicit rapidly fluctuating high-dimensional representations, encoding rich sensory information that has yet to be explained by state-of-the-art behavioral and computational models.

Keywords: neural dimensionality; temporal dynamics; visual processing; representation geometry; EEG

Introduction

How neural representations of visual information change over time remains a central question in neuroscience. Research demonstrates that various visual features can be decoded from neural time series data such as EEG and MEG (Grootswagers, Wardle, & Carlson, 2017; Contini, Wardle, & Carlson, 2017), showing that neural signals contain rich information about visual processing as it unfolds. Importantly, neural representations remain stable only within limited time frames (Isik, Meyers, Leibo, & Poggio, 2014; Carlson, Tovar, Alink, & Kriegeskorte, 2013), suggesting dynamic visual coding in the brain.

Despite our ability to decode visual features from neural signals, we lack an understanding of what properties of neural activity determine visual processing dynamics. Most studies focus on when information can be decoded (Bankson, Hebart, Groen, & Baker, 2018; Teichmann, Hebart, & Baker, 2024) rather than characterizing how neural representations evolve in their geometric properties. Recent theoretical frameworks suggest neural dimensionality constrains computational capacity (Stringer, Pachitariu, Steinmetz, Carandini, & Harris, 2019; Gauthaman, Ménard, & Bonner, 2024), but few studies have examined how dimensionality relates to temporal decoding patterns in visual processing.

Here, we investigated EEG data dimensionality from the THINGS-EEG2 dataset (Gifford, Dwivedi, Roig, & Cichy, 2022) in relation to decoding both behavioral features (Hebart et al., 2023) and visual neural network representations (Radford et al., 2021). Our findings show that the dimensionality of EEG signals rapidly increases following stimulus onset and tracks the decodability of both behavioral and DNN embeddings over time. We find that previously documented patterns of above- and below-chance accuracies in temporal-generalization analyses (King & Dehaene, 2014; Carlson et al., 2013) are associated with the shared dimensionality of neural representations across time points. Notably, the shared space between neural data and features (both behavioral and model-based) has substantially lower dimensionality than the EEG signal itself, revealing aspects of neural activity that elude current models of visual-feature representation.

Methods

Time series dataset We analyzed the THINGS-EEG2 dataset (Gifford et al., 2022), comprising EEG recordings from 10 participants viewing objects from the THINGS database (Hebart et al., 2023). We focused on 17 channels from occipital and parietal cortex. The train set includes responses to 16,540 unique images (4 repetitions each), and the test set includes 200 images (80 repetitions each).

Features of interest We examined two feature types: 66 behavioral features from the THINGS dataset (Hebart et al., 2023) derived from a triplet odd-one-out behavioral task, and model features from the first and last Identity layers of Open-CLIP ResNet50 (Radford et al., 2021; Ilharco et al., 2021). Following Conwell, Prince, Kay, Alvarez, and Konkle (2024), we determined the number of projections via the Johnson-Lindenstrauss lemma and applied sparse random projection to the model activations, resulting in 8,336 dimensions.

Linear mapping and temporal generalization We trained ridge regression ($\alpha = .01$) to predict features from neural data in the train set, computing Pearson correlations between predicted and actual values in the test set. Neural data were averaged across repetitions. For temporal generalization, we applied decoder weights from each time point to all other time points.

Dimensionality For EEG dimensionality, we fit PCA on training neural data, split the test data into random halves, and computed Pearson correlations between these projections (Stringer et al., 2019; Gauthaman et al., 2024). For shared space dimensionality, we fit PLSSVD under the same procedure as regression and computed correlations between projections of test data and features onto the shared space. Given the correlation results from the above analyses, we conducted cluster-based nonparametric tests (Maris & Oostenveld, 2007) for each dimension across all times, calculating positive dimensionality as the number of dimensions significantly above null, and negative dimensionality as those significantly below null.

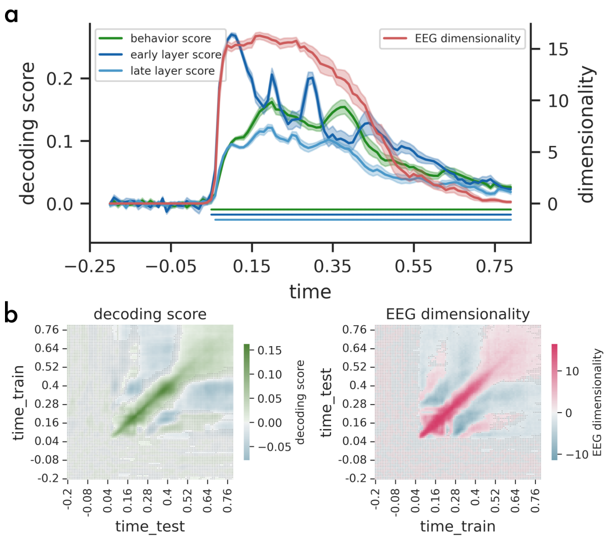

Figure 1: **a)** Time course of decoding for behavioral features and early and late ResNet50 layers compared with EEG dimensionality. Solid lines underneath show significant above-chance decoding periods, which align with high EEG dimensionality. Shaded areas represent standard error of mean across subjects. **b)** Temporal generalization matrices for feature decoding (left) and EEG latent dimensions (right) reveal a correspondence between decoding accuracy and shared dimensionality patterns. Oscillations of neural latent dimensions give rise to above- and below-chance accuracies for feature decoding. Shaded cells indicate non-significant values.

Results & Discussion

Our analyses reveal a relationship between neural dimensionality and the temporal dynamics of visual representations. As shown in Figure 1a, EEG dimensionality increases sharply after stimulus onset, peaking between 150-350ms before gradually declining. Notably, the EEG data approaches full rank during this peak period despite the high signal spread across channels and low signal-to-noise ratio that are typically inherent to EEG measurements (Michel & Brunet, 2019), suggesting remarkably structured neural responses to visual stimuli. This profile closely tracks periods of significant decoding of both behavioral and model features from EEG data, suggesting an association between neural geometric structure and feature decodability.

The temporal generalization matrices (Figure 1b) demonstrate time-specific neural representations, with shared dimensionality tracking significant above-chance decoding scores. Interestingly, previous studies have reported significantly below-chance generalization across certain time points (King & Dehaene, 2014; Carlson et al., 2013), which we also observe. By computing negative shared dimensionality, we characterize these representational inversions up to full rank, rather than only the target features.

When comparing the dimensionality of EEG data with that of shared spaces between EEG and visual features (Figure 2),

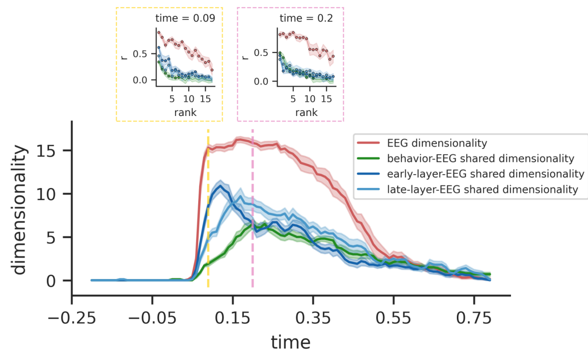

Figure 2: EEG dimensionality compared to shared-space dimensionality between EEG and various visual features. The consistently lower shared-space dimensionality indicates that both behavioral and model features capture only a subset of the high-dimensional neural signal. Shaded areas represent standard error of the mean across subjects. Insets show correlation by rank at two time points , with markers indicating dimensions significantly above chance.

EEG data maintains up to 15 significant dimensions on average during peak processing, while the shared space with behavioral features reaches only about 7 dimensions. The shared space with model features captures more dimensions than behavioral features, with early features better aligning with initial processing and late features with sustained processing. These model representations, however, still account for only a portion of the EEG dimensionality, which is full rank. This suggests that representational dynamics of visual content extend beyond what can be characterized through behavioral or model-based features, pointing to aspects of neural activity not captured by our current understanding of visual information processing.

Our findings highlight dimensionality as an important aspect of neural information processing. By computing the geometric properties of temporal generalization, we provide a framework that reveals patterns in how neural codes evolve over time. Critically, the substantial unexplained dimensions we've identified in EEG signals—beyond both behavioral features and computational models—suggest that neural activity contains rich dynamic information beyond what current feature-based decoding analyses reveal.

References

Bankson, B. B., Hebart, M. N., Groen, I. I., & Baker, C. I. (2018). The temporal evolution of conceptual object representations revealed through models of behavior, semantics and deep neural networks. *NeuroImage*, *178*, 172–182.

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: the first 1000 ms. *Journal of vision*, *13*(10), 1–1.

Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia*, *105*, 165–176.

Conwell, C., Prince, J. S., Kay, K. N., Alvarez, G. A., & Konkle, T. (2024). A large-scale examination of inductive biases shaping high-level visual representation in brains and machines. *Nature communications*, *15*(1), 9383.

Gauthaman, R. M., Ménard, B., & Bonner, M. F. (2024). Universal scale-free representations in human visual cortex. *arXiv preprint arXiv:2409.06843*.

Gifford, A. T., Dwivedi, K., Roig, G., & Cichy, R. M. (2022). A large and rich eeg dataset for modeling human visual object recognition. *NeuroImage*, *264*, 119754.

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. *Journal of cognitive neuroscience*, *29*(4), 677–697.

Hebart, M. N., Contier, O., Teichmann, L., Rockter, A. H., Zheng, C. Y., Kidder, A., . . . Baker, C. I. (2023). Things-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. *Elife*, *12*, e82580.

Ilharco, G., Wortsman, M., Wightman, R., Gordon, C., Carlini, N., Taori, R., . . . Schmidt, L. (2021, July). *Openclip*. Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.5143773` doi: 10.5281/zenodo.5143773

Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. *Journal of neurophysiology*, *111*(1), 91–102.

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in cognitive sciences*, *18*(4), 203–210.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of eeg-and meg-data. *Journal of neuroscience methods*, *164*(1), 177–190.

Michel, C. M., & Brunet, D. (2019). Eeg source imaging: a practical review of the analysis steps. *Frontiers in neurology*, *10*, 325.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., . . . others (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748–8763).

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., & Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature*, *571*(7765), 361–365.

Teichmann, L., Hebart, M. N., & Baker, C. I. (2024). Dynamic representation of multidimensional object properties in the human brain. *bioRxiv*, 2023–09.