

# Spatio-temporal Dynamics of Visual Object Recognition in the Human Brain

Yuvraj S. Sethi

Phillips Exeter Academy, 20 Main St., Exeter, NH, 03833, U.S.A.; ysethi@exeter.edu

**ABSTRACT:** The human brain is fundamentally a big data processor in terms of both image content and neural data. Yet, magnetoencephalography (MEG) neuroimaging studies are typically conducted with small numbers of visual stimuli, which may not support generalizations because they severely under-sample the stimulus space. This study analyzes available MEG data on human brain response across a large stimulus space spanning images of indoor and outdoor scenes, multiple objects, and singular objects. The images were divided into four categories: faces vs objects; large scenes vs small scenes; multiple objects vs single object; and moving (action) vs static. Time-resolved decoding was performed using a linear support vector machine classifier to estimate the time series with which categorical content emerges in the human brain. Decoding results were robust, reaching 100% accuracy as early as 100-130ms from the onset of the stimuli for all categories, excluding the action vs. static category that yielded relatively weaker decoding results. Overall, the results indicate that decoding several visual categorical representations with MEG data is possible even with very large numbers of diverse naturalistic image stimuli. The findings pave the way to future studies for exploring the critical dimensions of scene processing in the human brain.

**KEYWORDS:** Biology; Neuroimaging; Neural Patterns; Magnetoencephalography; Multivariate Pattern Analysis.

## ■ Introduction

The recognition of complex visual patterns emerges from neural activity in a distributed network of cortical regions. In studying the primary visual cortex, an important question in cognitive neuroscience arises: how does the brain manage complex tasks such as instantly recognizing and distinguishing a familiar object? In recent years, the field of neuroscience has been transforming: Multivariate Pattern Analysis (MVPA) tools are enabling technologies such as Magnetoencephalography to decipher neural patterns and begin to formulate a comprehensive understanding of the complexity of the human brain. Previously, progress towards this goal has been hindered by the limitation of non-invasive brain imaging technologies and their inability to simultaneously decode temporal and spatial information.<sup>1</sup>

Unlike functional magnetic resonance imaging (fMRI), electroencephalography (EEG), positron emission tomography (PET), and single-photon emission computed tomography (SPECT), magnetoencephalography (MEG) provides a direct measure of brain function rather than functional measures and has very high temporal and spatial resolution. MEG is a non-invasive neurophysiological technique that measures the magnetic fields generated by neuronal activity of the brain. The nature of this technique means the source of an event can be localized with millimeter precision and the timing with millisecond precision.<sup>2</sup>

The introduction of multivariate decoding methods has made a significant impact on advancing cognitive neuroscience. Multivariate pattern analysis of MEG data provides detailed insights into the temporal dynamics of neural activity, containing information about cognitive and sensory process-

es.<sup>1</sup> This decoding of brain activity can help assess not only cognitive function but also provide a basis for comparing normal and abnormal function of the brain. These classification methods have been used to study simple visual features not limited to complex visual patterns, and auditory functions of the brain.<sup>3</sup> Further research into these methods could lead to a deeper understanding of the biological basis of several brain disorders and aid the development of imaging biomarkers leading to clinical treatments.

Decoding visual categorical information in MEG with high temporal and spatial information can provide insights into how categorical content emerges in the human brain, leading to a deeper understanding of temporal dynamics of cognitive representations. The scope of this study is to decode information captured with MEG from evoked responses using a very large number of stimuli for brain activity over a period of time. Sample stimuli is shown in Figure 1. Associating and distinguishing the brain patterns based on the corresponding stimulus can help researchers determine if valuable information on brain representations can be derived from the MEG neuroimaging data.<sup>4</sup>

To analyze the time course of brain processing for applied stimuli, multivariate pattern analysis was performed on the MEG data in a time-resolved manner. A strength in doing such time-series decoding is that the dynamic changes in brain patterns can be examined across multiple time points yielding a time series that accurately describes how the brain processes information. Figure 2 below depicts the multivariate pattern analysis pipeline.

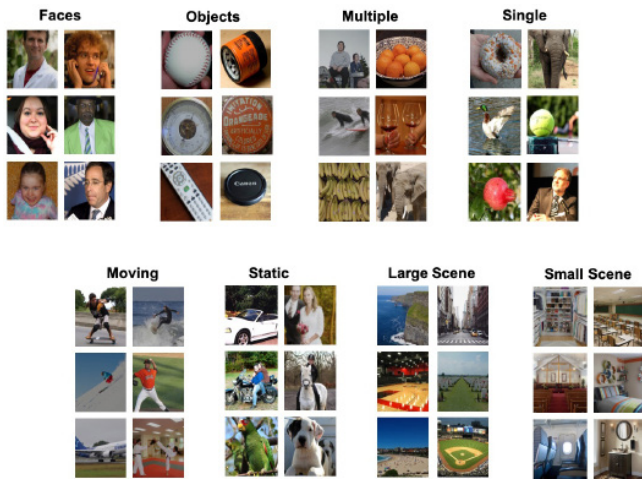


Figure 1. Sample images across multiple categories used as stimuli to evoke response and capture neural patterns.

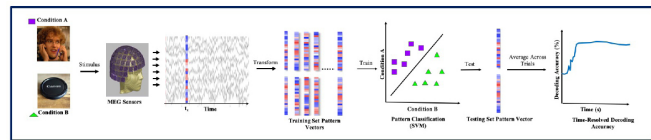


Figure 2. Multivariate pattern analysis pipeline.

Previous studies for analyzing MEG data for visual recognition have generally employed a small stimulus space.<sup>3</sup> This study is unique in its investigation of decoding across a large number of stimuli which de-emphasizes the impact of low-level image features, such as pixel intensity, pixel gradient orientation, and color, and puts more focus on high-level semantic image content such as objects and actions in order to deepen our understanding of information processing for semantic visual recognition.

### Results and Discussion

This study explored the spatio-temporal dynamics of cognitive processes using a large stimulus space. The stimuli represented a set of 4,916 unique images, which were divided into eight subcategories: faces (438 images), objects (547 images), multiple (874 images), single (907 images), moving (669 images), static (622 images), large scenes (1004 images), and small scenes (944 images). The data was recorded using MEG and was analyzed using multivariate pattern classification methods (time-resolved decoding using a linear support vector machine classifier). The use of 306 MEG sensors allowed for both millisecond and millimeter precision in analyzing how categorical content emerges in the human brain.

Data was recorded from -200 ms to +700 ms after stimulus onset. At each time point (tx), a linear SVM classifier was used for decoding each of the four categories (faces vs objects, multiple vs single images, moving vs static images, and large scenes vs small scenes). Figure 3 shows the time series representation of neural pattern evolution in the brain. Figure 4 highlights the difference in brain response to different stimuli on a time scale.

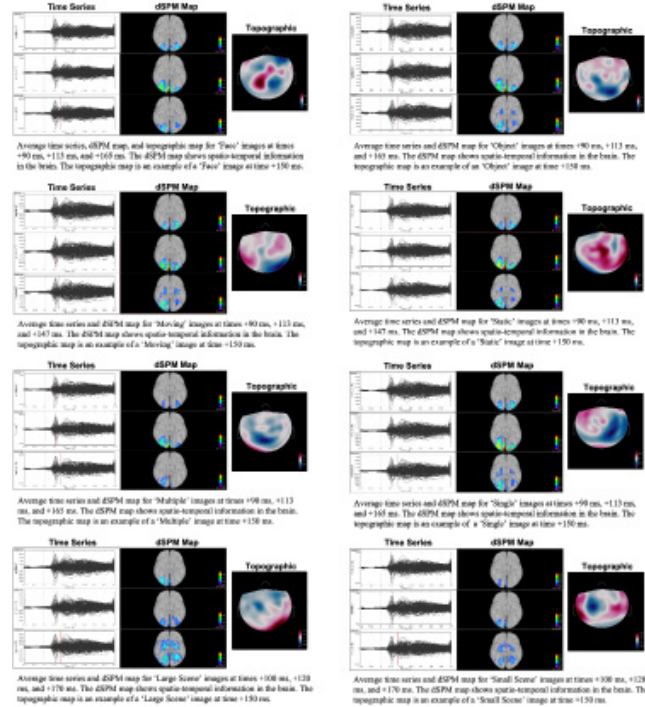


Figure 3. Time series representation of neural pattern evolution.

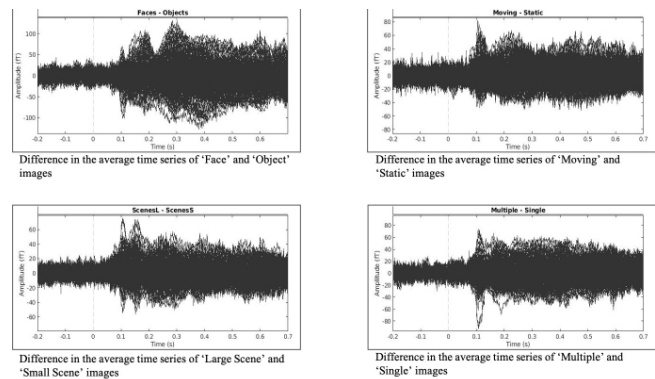


Figure 4. Difference in time series of evoked responses.

As shown in Figure 5, decoding results were robust; they reached 100% accuracy as early as 100-150ms from the onset of the stimuli for all categories, excluding the moving vs. static images category. Decoding time series for faces vs objects fluctuated between 80% and 95% after stimulus onset until +150 ms, thereafter remaining at 100% for an extended period of time until after +700 ms. Decoding time series for multiple vs single reached 100% decoding accuracy as early as +105 ms, with little fluctuation until +700 ms after stimulus onset. Large scenes vs small scenes similarly achieved 100% decoding accuracy as early as +105 ms, however there was marginally greater fluctuation between 90% and 100% accuracy until +250 ms, subsequently remaining at ~100% accuracy. The moving vs static image category, which yielded relatively weaker decoding results, had a decoding accuracy varying between 78% and 100% (78% at +150 ms) from +100 ms until +700 ms after stimulus onset. This was to be expected because of the similarities between a “static” image and a “moving” image which

captures a moment of action. In comparison, the other three categories had far more defined boundaries differentiating them from one another.

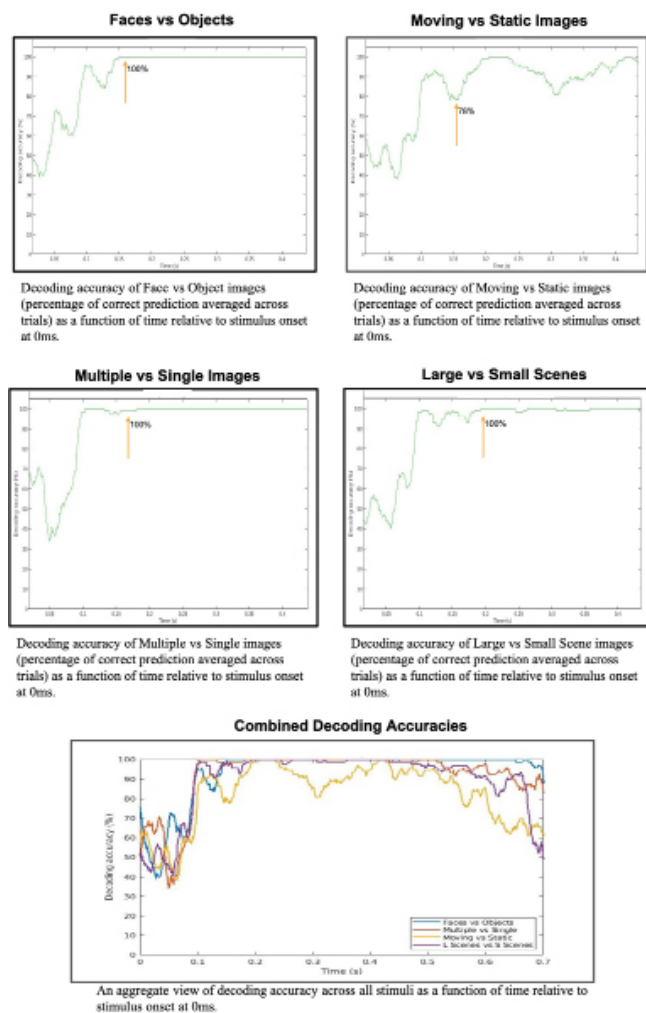


Figure 5. Multivariate classifier decoding accuracy.

## Conclusion

Using the linear support vector machine classifier, we were able to decode the MEG sensor data which is critical in understanding how brain response varies across different image stimuli. Successfully using this classifier to identify these responses opens up the opportunity to apply this research to auditory, olfactory, tactile, and gustatory stimuli. The research shows us that MEG data can effectively capture subtle variations in brain response with both high spatial and temporal resolution. By leveraging the classifier, we will be able to derive meaningful outcomes from the MEG data to advance our knowledge of information processing in the human brain.

There are multiple opportunities to enhance the study further, such as extending the work to include temporal generalization. Temporal generalization of decoding analysis enables us to distinguish transient brain patterns from persistent brain patterns. The decoding procedure is very similar to the one used in this study; however, one could train the SVM classifier at a given

time point and test it across all other time points rather than the given time point. If the brain pattern representations are consistent over time, it implies that for a given stimulus, the neural activity is similar over that time interval. Similarly, this approach can also be used to test cognitive models which can help us understand the association of brain response between two completely different stimuli, where the SVM classifier is trained on one stimulus (ex. faces) but tested on different stimuli (ex. objects).

It must be noted that decoding can only confirm that a neural pattern corresponds to a given stimuli and therefore contains “some” information but does not prove this information is used by the brain. One approach to determine the functional relevance of decoding results would be to analyze the brain patterns triggered from a behavioral response, for example, correlating the decoding accuracy with reaction time (ex. pressing a button).

## Methods

The study utilized the raw MEG data originally collected across ten sessions based on a large stimulus set.<sup>5</sup> The data was made available for this research. The dataset consisted of brain response to 4,916 unique images which were separated into eight different subcategories: faces, objects, multiple, single, moving (action), static, large scenes, and small scenes. The categorized data was analyzed using Brainstorm, an open-source software dedicated to MEG and EEG data analysis, written in primarily Matlab scripts.<sup>6</sup> For every category of data, an arithmetic average was calculated, resulting in an average time series, Dynamic Statistical Parametric Map (dSPM), and MEG topography. The difference between the average time series for each of the four categories, resulted in a new time series showing variation in signals between subcategories.

Next, decoding was performed using a linear Support Vector Machine (SVM) classifier. At time  $t_x$ , using k-fold cross-validation (Figure 6), the data was divided (ex. Faces and Objects) five-fold (subsets) to be used for classification, where each of the five subsets consisted of two 306 pattern vectors: one of faces and one of objects. Each set contained a proportionate number of face and object stimuli (one-fifth of the total faces and one-fifth of the total objects). Four of the subsets were used for the training set and the remaining subset as the testing set.

The process was repeated fifty times for the same  $t_x$ , creating a new set of folds for each permutation, where  $t$  ranged from -200 ms to 700 ms, in increments of 1 ms. By averaging the results of fifty trials (testing sets) at each time  $t_x$ , a decoding matrix was created resulting in the decoding time series (decoding accuracy vs time) for faces vs objects. This process was repeated for the remaining three categories: multiple vs single, moving vs static, and large scenes vs small scenes.

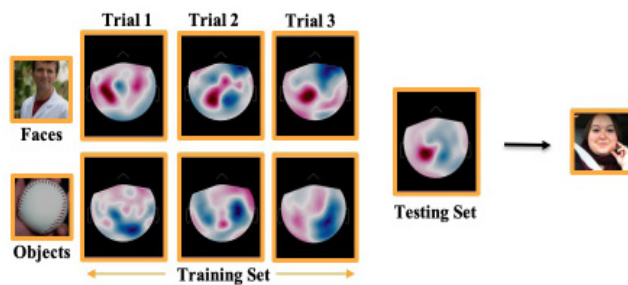


Figure 6. K-fold cross-validation procedure.

## ■ Acknowledgements

The author would like to thank Dr. Dimitrios Pantazis, McGovern Institute for Brain Research, MIT, for being a mentor for this research project and providing access to the MEG data and software tools to help accomplish the objective of this research.

## ■ References

1. Baillet, S. Magnetoencephalography for brain electrophysiology and imaging. *Nat Neurosci* 20, 327–339 (2017). <https://doi.org/10.1038/nn.4504>.
2. Chang, N., J.A Pyles, A. Marcus, A. Gupta, M. Tarr, and E. Aminoff. BOLD5000, a public fMRI dataset while viewing 5000 visual images. *Sci Data* 6, 49 (2019). <https://doi.org/10.1038/s41597-019-0052-3>.
3. Radoslaw MC., A. Khosla, P. Dimitrios, and O. Aude. “Dynamics of Scene Representations in the Human Brain Revealed by Magnetoencephalography and Deep Neural Networks.” *NeuroImage*, vol. 153, 2017, pp. 346–358., DOI:10.1016/j.neuroimage.2016.03.063.
4. Dobs, K., L. Isik, D. Pantazis, and N. Kanwisher. “How Face Perception Unfolds over Time.” *Nature Communications*, vol. 10, no. 1, 2019, DOI:10.1038/s41467-019-09239-1.
5. Gross, J., S. Baillet, G. Barnes, R. Henson, A. Hillebrand, O. Jensen, K. Jerbi, V. Litvak, B. Maess, R. Oostenveld, L. Parkkonen, J. Taylor, V. Wassenhove, M. Wibral, and J.M. Schoffelen. “Good practice for conducting and reporting MEG research.” *NeuroImage* vol. 65,100 (2013): 349–63. doi:10.1016/j.neuroimage.2012.10.001
6. Tadel, F., S. Baillet, J. Mosher, D. Pantazis, and R.M. Leahy. “Brainstorm: A User-Friendly Application for MEG/EEG Analysis.” *Computational Intelligence and Neuroscience*, vol. 2011, 2011, pp. 1–13., DOI:10.1155/2011/879716.

## ■ Authors

Yuvraj Sethi is a senior at Phillips Exeter Academy. He has a passion for scientific research in the field of healthcare. His research work has been published in IEEE conferences proceedings and Vision Science Society conferences. In his free time, Yuvraj enjoys playing football, the saxophone, and the piano.