# Representational dynamics of visual contents underlying object selectivity

**Chun-Hui Li (iamscli.tw@gmail.com)**
Freie Universität Berlin, Germany; National Taiwan University, Taiwan.

## Abstract

**Attention allows humans to prioritize relevant information in line with behavioral goals, yet the temporal dynamics of its influence on object selectivity remain unclear. In this study, we integrated magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and deep neural networks (DNNs) to investigate how goal-directed attention selectively modulates visual object representations. Representational similarity analysis (RSA) revealed enhanced category-specific similarity in brain regions under attention, with top-down modulation from prefrontal cortex to category-specific areas through Granger causality analysis. GLM-based RSA confirmed dynamic object-selective representations in the fusiform face area (FFA) and parahippocampal place area (PPA). Furthermore, MEG-DNN fusion showed attentional effects across hierarchical DNN layers, reflecting a coarse-to-fine processing pattern. These findings offer new insights into the spatiotemporal dynamics of object-based attentional modulation in the human brain.**

**Keywords:** Representational similarity analysis; MEG-fMRI-DNN fusion method; attention; face perception; scene perception;

## Introduction

Attention is essential for humans to prioritize and select relevant information from both the external environment and internal representations. This selective processing is guided by current behavioral goals (Van Ede & Nobre, 2023). Studies on visual attention have revealed psychological and neural mechanisms underlying top-down attentional modulation of object processing (Baldauf & Desimone, 2014; Mehrpour et al., 2020; Peelen & Kastner, 2014; Quek et al., 2018). However, the spatiotemporal dynamics of attentional modulation on object selectivity in the human brain remain largely unexplored, as does the extent to which goal-directed attention shapes the visual content that supports object selectivity.

To address these questions, we combined magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and deep neural networks (DNNs). We employed a fusion approach that links MEG sensor activation patterns with either neural activity patterns captured by fMRI or hierarchical feature representations derived from layers of AlexNet (Krizhevsky et al., 2012). Furthermore, we examined whether and how object-based attention dynamically modulates both coarse-to-fine and category-related information.

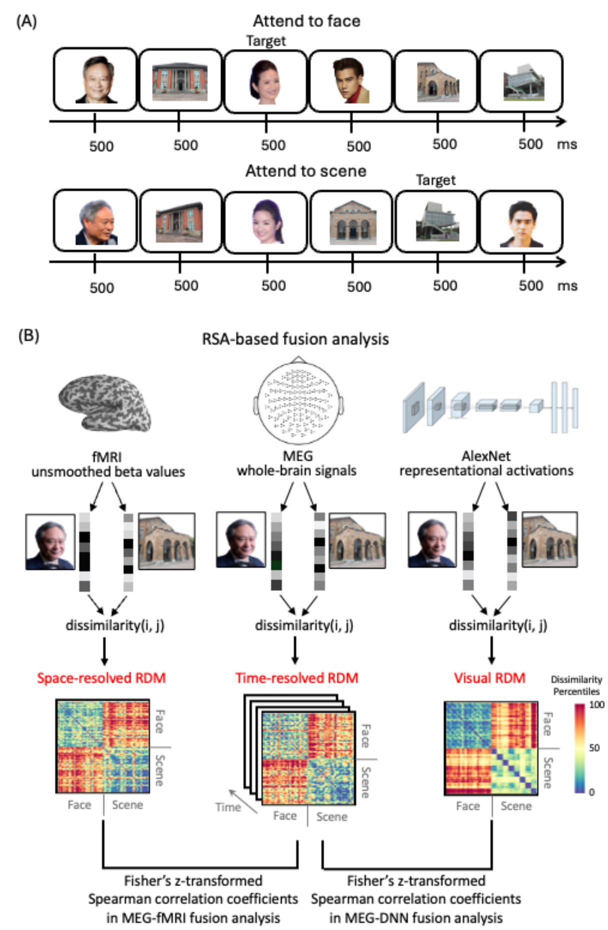

Figure 1: (A) Schematic illustration of the target detection task. (B) RSA-based fusion approach.

## Methods

Twenty-four participants took part in both MEG and fMRI experiments involving streams of face and scene images (Fig. 1A). MEG data were recorded using a 306-channel Elekta Neuromag Triux system, and fMRI data were acquired on a Siemens 3T Prisma scanner. During the MEG session, participants were instructed to attend to either faces or scenes and respond to predefined targets within the attended category. In the fMRI session, they completed a visual localizer task using the same face and scene images, along with their scrambled counterparts, presented in a block design.

### Representational Similarity Analysis

We performed RSA-based fusion analyses to integrate MEG data with either fMRI or DNN data (Fig. 1B) (Cichy et al., 2014, 2016). Regions of interest (ROIs), including the fusiform face area (FFA) and parahippocampal place area (PPA), were defined using fMRI contrasts and the Automated Anatomical Labeling (AAL) template. For each modality, pairwise dissimilarities were computed across all stimulus and condition combinations, resulting in representational dissimilarity matrices (RDMs).

Each RDM, symmetrical along the diagonal, was reduced to a vector by extracting its lower triangular part. For each ROI, we computed Spearman correlation coefficients between MEG and fMRI RDMs, separately for face-selective and scene-selective regions, to characterize the temporal evolution of representational similarity. Similarly, we computed Spearman correlations between MEG and DNN RDMs, assessing representational alignment under different attention conditions.

Finally, we also constructed a linear model using face and scene RDMs as regressors to predict the time-resolved MEG RDMs, enabling a fine-grained investigation of the neural dynamics underlying object-based attentional modulation.

## Results

The fusion analyses revealed that attention enhanced representational similarity within category-selective brain regions. Granger causality analysis further indicated that this goal-directed modulation was influenced by the prefrontal cortex. Crucially, GLM-based RSA confirmed the spatiotemporal dynamics of object-based attention within these category-selective areas (Fig. 2A).

In addition, MEG-DNN fusion showed similar attentional effects on representational similarity across hierarchical layers of AlexNet (Fig. 2B). We also observed a significant increase in attentional modulation similarity from shallow to deep layers (Fig. 2C). These findings demonstrate how, when, and to what extent the visual representations within the two distinct object-selective brain areas are selectively modulated by top-down attention.

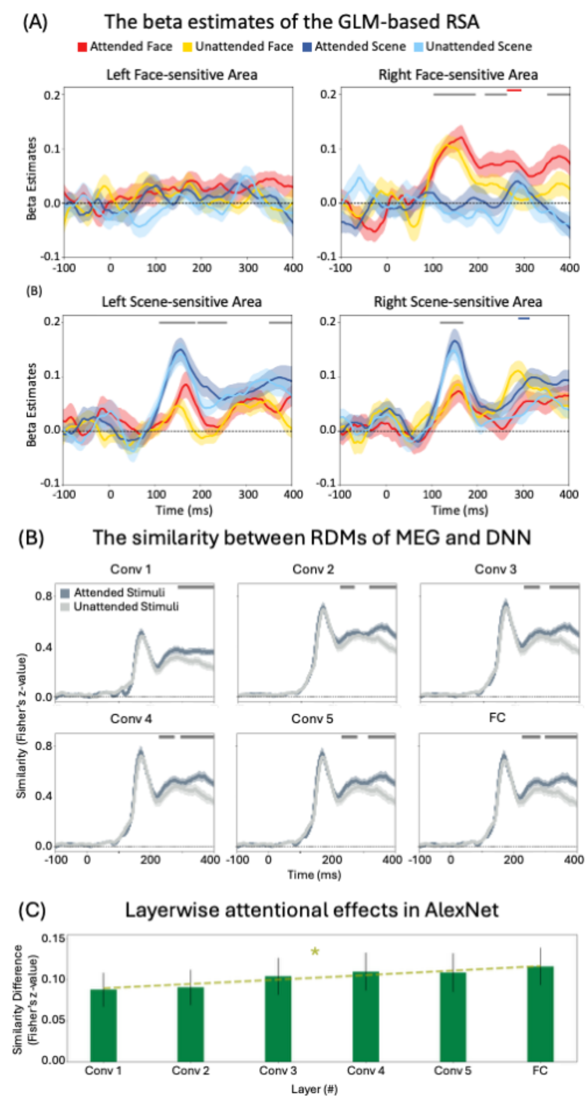


Figure 2: (A) Beta estimate results of the GLM-based MEG-fMRI RSA. (B) Similarity results of MEG-DNN RSA. (C) Layerwise attentional effects in AlexNet.

## Acknowledgements

This work was supported by National Science and Technology Council, Taiwan (111-2424-H-002-012-DR, 113-2811-H-002-006, 113-2423-H-008-004-). The author would like to express sincere gratitude to Dr. Bo-Cheng Kuo for his invaluable guidance, insightful feedback, and unwavering support throughout the course of this research.

## References

Baldauf, D., & Desimone, R. (2014). Neural Mechanisms of Object-Based Attention. Science, 344(6182), 424–427. https://doi.org/10.1126/science.1247003

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. Scientific Reports, 6(1), 27755.

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

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25. https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

Mehrpour, V., Martinez-Trujillo, J. C., & Treue, S. (2020). Attention amplifies neural representations of changes in sensory input at the expense of perceptual accuracy. Nature Communications, 11(1), 2128.

Peelen, M. V., & Kastner, S. (2014). Attention in the real world: Toward understanding its neural basis. Trends in Cognitive Sciences, 18(5), 242–250.

Quek, G., Nemrodov, D., Rossion, B., & Liu-Shuang, J. (2018). Selective attention to faces in a rapid visual stream: Hemispheric differences in enhancement and suppression of category-selective neural activity. Journal of Cognitive Neuroscience, 30(3), 393–410.

Van Ede, F., & Nobre, A. C. (2023). Turning Attention Inside Out: How Working Memory Serves Behavior. Annual Review of Psychology, 74(1), 137–165. https://doi.org/10.1146/annurev-psych-021422-041757