Similarity-based fusion of MEG and fMRI discerns early feedforward and feedback processing in the ventral stream

Yalda Mohsenzadeh[1], Radoslaw M. Cichy[2], Aude Oliva[1], Dimitrios Pantazis[1]

[1] Massachusetts Institute of Technology, Cambridge, MA, USA
[2] Free University Berlin, Berlin, Germany

Computational models of vision have become ubiquitous in many domains of science, with the most powerful models now competing with human performance in several tasks. Many popular models, such as deep neural networks and HMAX, are inspired by the human visual system, relying on a hierarchical cascade of feedforward transformations akin to the human ventral stream. Despite these advances, the human visual cortex remains unique in complexity, with feedforward and feedback pathways characterized by rapid spatiotemporal dynamics as visual information is transformed into semantic content. Thus, a systematic characterization of the spatiotemporal and representational space of the ventral visual pathway can offer novel insights in the duration and sequencing of cognitive processes, suggesting computational constraints and new architectures for computer vision models.

To discern the feedforward and feedback neural processes underlying human vision, we used MEG/fMRI fusion (Cichy et al 2014, 2016). We collected MEG data while observers (N =17) viewed a rapid serial visual presentation (RSVP) of 11 images with an extremely fast speed (17ms/picture or 34ms/picture). Participants performed a two-alternative forced choice task reporting whether the middle image is a face or non-face. fMRI data while observers viewed the same stimuli were also collected.

We used multivariate pattern analysis to pairwise compare all stimuli, creating representational dissimilarity matrices (RDMs) separately for MEG and fMRI data. Comparison of time-resolved MEG RDMs with space-resolved fMRI RDMs yielded a spatiotemporal description of the ventral stream dynamics (Figure 1). Starting from early visual cortex (EVC), brain activation progressed rapidly to IT within approximately 110 ms from stimulus onset. The activation cascade reversed back to EVC at around 170 ms. This reversal was accompanied by a strengthening of IT activation, leading to categorical representations sustained over hundreds of milliseconds.

Taken together, our results revealed a clear dissociation between feedforward and feedback early visual processes, with well-defined temporal signatures for both mechanisms. The spatiotemporal dynamics can be used as constraints for developing new computational neuroscience models with recursive processes, to increase performances in challenging visual processing conditions.

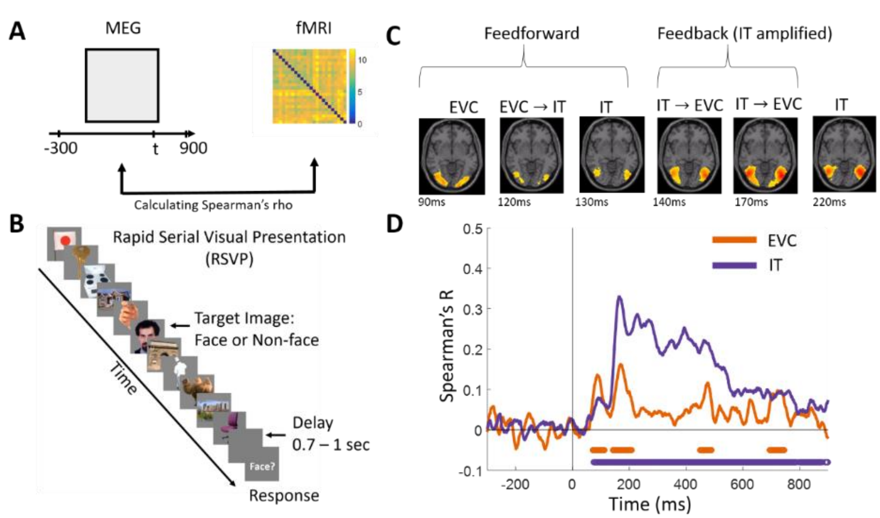

Figure 1. Fusion of MEG-fMRI data dissociates feedforward from feedback processes. **A)** Time-resolved MEG representational matrices were compared (Spearman's rho) with space-resolved fMRI representational matrices, resulting in fine description of visual brain dynamics. **B)** Rapid serial visual presentation of images. Participants viewed a sequence of 11 images presented with the speed of 34ms/pic or 17ms/pic, and then responded by pressing a button whether they saw a face in the stream. **C)** Frames of a movie showing the time course of feedforward and feedback flow of information in the brain. **D)** ROI-based analysis of the same MEG-fMRI fusion data for two representative brain areas, EVC and IT.