The AAAI 2017 Spring Symposium on
Science of Intelligence: Computational Principles of Natural and Artificial Intelligence
Technical Report SS-17-07

Content-Dependent Fusion: Combining Human MEG and FMRI Data to Reveal Spatiotemporal Dynamics of Animacy and Real-world Object Size

Seyed-Mahdi Khaligh-Razavi*[1], Radoslaw M. Cichy[2], Dimitrios Pantazis[3], Aude Oliva[1]

[1]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA

[2]Department of Education and Psychology, Free University Berlin, Berlin, Germany.

[3]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

*corresponding author: skhaligh@mit.edu

Abstract

Understanding the computational principles used by the human brain to perform recognition tasks requires a comprehensive view of which and when different brain regions come together to make sense of sensory information. Here we extend the neural fusion approach of Cichy et al (2016), which combines high spatial resolution (fMRI) and high temporal resolution (MEG) brain data using Representational Similarity Analysis (RSA), to selectively track the spatio-temporal dynamics of two major cognitive dimensions of objects: animacy and real-world size. The content-dependent fusion method proposed here determines when brain regions involved in object processing have a representation of animacy or size. We find that around ~160ms after image onset, the right lateral-occipital (LO) and posterior inferior-temporal (IT) regions encode a representation of animacy, whereas a representation of the real-world size of objects originates from left LO and bilateral parahippocampal (PHC) regions a little later, at ~168ms. Analysis of the content-dependent fusion dynamics shows that animacy representation is quite sustained over time, whereas real-world size representation is transient. Deciphering the sequencing of information processing and its persistence nature offer new architectural insights for computational models of recognition.

Content-dependent fusion

Recent developments in neuroscience data mining techniques (Hutchison et al. 2013, Cichy et al. 2014, 2016; Khaligh-Razavi & Kriegeskorte 2014, Yamins & DiCarlo 2016) have given the tools to study how separate brain regions work together in networks. Furthermore, by relating MEG and fMRI data, we can obtain time-stamps of spatial maps observed in fMRI. Cichy et al. (2016) used representational similarity analysis (RSA) (Kriegeskorte & Kievit 2013) to relate these two modalities. By comparing similarity relations between MEG patterns overtime to fMRI patterns over voxels, they were able to calculate a spatio-temporally resolved view of the emergence of visual representation in the brain, termed fusion. Here we propose a *content-dependent fusion* (Fig 1) that similarly uses RSA to related MEG and fMRI representations; but further enables us to distinguish the type of visual information (i.e. content) that is being carried from one brain area to another, and when. Simply, the content-dependent fusion can be thought of as an MEG-fMRI similarity based fusion (Cichy et al, 2016) that is further confined by a constraint on the explained similarity variance between MEG and fMRI. The new method also allows us to visualize the representational connectivity finger prints of different kinds of information (here object animacy, and real-world size) over time.

MEG and fMRI data: We used the MEG and fMRI data from Cichy et al. (2016). Fifteen subjects were presented with 118 object stimuli for 0.5 second each in independent MEG and fMRI sessions. For our purpose, each of the 118 images were further classified by their animacy (animate/inanimate) or by their real-world size (small: hand-size / large: body-size). See Cichy et al. (2016) for the full set of stimuli and method description.

Content-Dependent Fusion: The proposed method is illustrated in Fig 1 with the steps and statistics further explained below:

1) Specify the representational dissimilarity matrix (RDM) for your reference model RDM (e.g. animacy model RDM). This is the content (or the type of information) whose spatio-temporal dynamics will be characterized in the brain.

2) Correlate the MEG RDMs with the reference model RDM, and find time-points with significant correlations (one-sided sign-rank test over N=15 subjects; FDR corrected across time at $p = 0.05$).

3) Perform a fMRI searchlight (5 voxel radius) correlating fMRI searchlight RDMs with the reference model RDM to

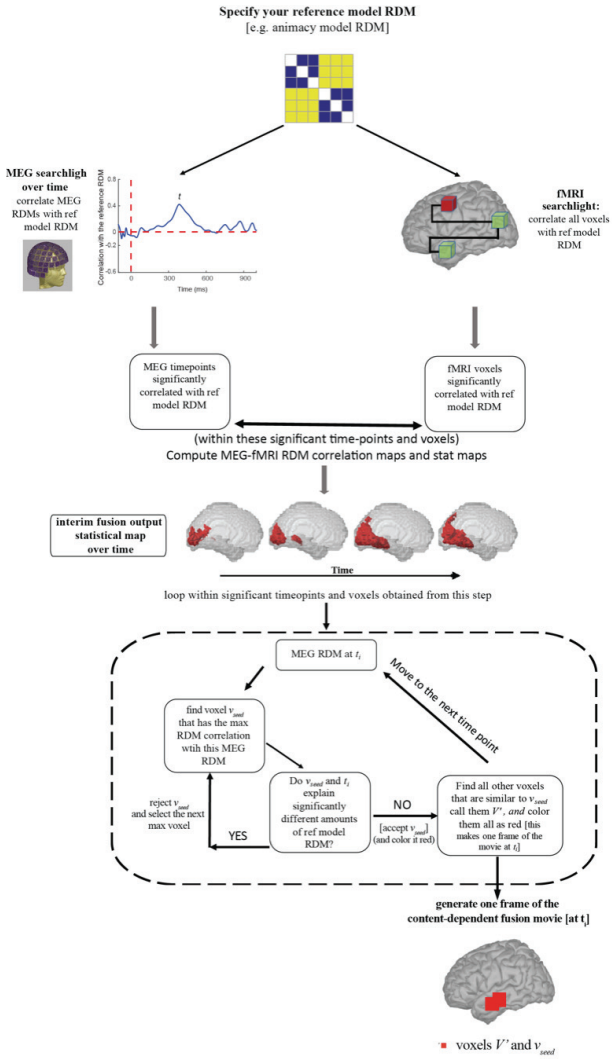

Fig 1. Content-dependent fusion stepwise algorithm: after specifying the reference model RDM (the content), two searchlights are performed: over time on the MEG data, and over space on the fMRI data. Then within significant time-points and voxels, each voxel is correlated with all the MEG RDMs over time, which gives us a correlation map over time and space. Using sign-rank test across subjects, statistical maps are obtained over these correlation maps (FDR corrected). This gives the interim fusion movie. Then we loop through voxels/time-points that remained significant and find corresponding fMRI voxels and MEG time-point that do not significantly differ in explaining the reference model RDM.

identify significant voxels (one-sided sign-rank test over 15 subjects; FDR corrected across voxels at p = 0.05).

4) Within MEG time-points and fMRI voxels that are significantly correlated with the reference model RDM, perform a fMRI-MEG fusion. That is, to correlate MEG representations at each time-point to fMRI representations at each voxel. This allows us to relate temporal dynamics observed in MEG with their spatial origins as indicated by fMRI (one-sided sign-rank test over N subjects; FDR corrected across voxels and time-points at p = 0.05).

5) Loop over MEG time-points that are significantly correlated with the reference model RDM (calculated in step 2). At each time-point take the MEG RDM as the reference and then follow step 6. [Next we will find voxels that correspond to this time-point and carry information about the reference model RDM.]

6) Loop over all voxels; among voxels whose RDMs are significantly correlated with both the reference model RDM and the reference MEG RDM, find the one that has the max correlation with the reference MEG RDM. This is the voxel that best relates to the reference MEG RDM, that is it best resembles the representation obtained at that time-point. Denote this as the 'seed voxel'.

7) Correspondency check: This will ensure that the corresponding MEG time-points and fMRI voxels do not have different representations with regard to the reference model RDM (i.e. they are similarly correlated with the reference model RDM, e.g. animacy model RDM). To this end, check if the reference model RDM explains a significantly different amount of variance of the seed voxel RDM compared to the reference MEG RDM (relative to the noise ceiling). If yes, this seed voxel is rejected, return to step 6, and take the next maximum point.

To do the statistical comparison for this step:

a) at time *t* and voxel *v*, identify all the 15 MEG and 15 fMRI subject RDMs

b) correlate the reference model RDM with all these 2N RDMs

c) then do a paired sign-rank test on the corrected correlations (corrected relative to the noise-ceiling) to see if these two are significantly different.

d) the seed voxel is rejected if they are significantly different, otherwise proceed to the next step.

Note about noise-ceiling: the noise-ceiling at each time-point and at each voxel is computed by correlating each subject's RDM with the subject-averaged RDM in that time-point/voxel. This gives us N correlations (one for each subject), which is then averaged and is used as an estimate for the noise-ceiling. (This provides an upper bound on the noise-ceiling as defined in Nili et al. (2014))

8) Define the selected seed voxel *'v'* and expand around it in the following way:

a) take all other voxels that are significantly correlated with the reference MEG RDM (already calculated at step 4) and pass the correspondency check (i.e. step 7), call them *V'*.

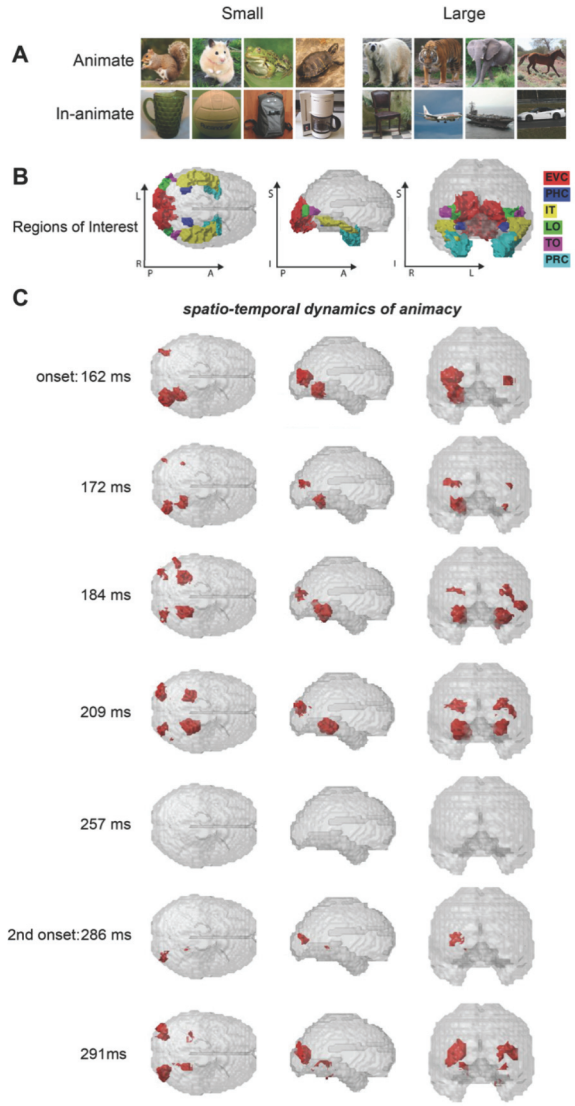

Fig 2. Spatiotemporal dynamics of animacy. A subsample of the stimulus set divided by animacy and real-world object size is shown (A). The original set contains 118 images in total. Content-dependent fusion analyses revealed neural dynamics of animacy across the whole brain over time (C). For reference, six regions of interest are shown (B). These regions are obtained from Wang et al. (2015) probabilistic atlas. Red voxels indicate statistical significance (N=15, one-sided sign-rank test, FDR-corrected at 0.05 across across time [-100,1101 ms] and space [all voxels]).

b) load all the N fMRI subject RDMs for each of these V' voxels.

c) for each voxel v_i in V' check whether the seed voxel RDM and the v_i RDM explain the reference MEG RDM differently, if yes then remove v_i from V'.

9) All the voxels remaining in V' after step 8, together with the seed voxel v, are the voxels corresponding to the reference MEG RDM at time-point t that carry information about the reference model RDM.

10) Repeat steps 5 to 9 to until we loop over all significant MEG time-points.

Results & Discussion

Content-dependent fusion enabled us to characterize and compare the spatio-temporal neural dynamics of semantic properties or classes, here animacy and object real-world size.

Animacy: Figure 2 shows that both left and right LO (lateral occipital) and posterior IT (inferior temporal) have a representation of the animacy information of objects at about 162ms after stimulus onset. Animacy is also captured by the PHC (parahippocampal cortex) and other higher-visual areas; with a right hemisphere dominance. We observe cycles of representation (animacy being on/off in these regions of interest) a behavior that needs further investigation to determine whether it characterizes recurrent processing.

These results are consistent with previous findings that animacy is strongly reflected in higher-stream visual areas, such as LO and IT (Kiani et al., 2007; Kriegeskorte et al., 2008; Naselaris et al., 2012). However, the fMRI studies could not reveal when each of these areas represented animacy. Similarly, MEG studies alone (Carlson et al., 2013) cannot locate the exact cortical areas involved in carrying animacy information. Here, combining both MEG and fMRI data within the content-dependent fusion enables us to answer both time and space questions.

Real-world object size: As shown in Figure 3, the spatio-temporal representation of real-world object size is different from animacy representation: size representation starts from left LO and bilateral PHC at ~168ms after stimulus onset. At 190 ms, the right TO (temporal occipital) has a representation of size as well. These regions are consistent with the regions found by Konkle and Oliva (2012), with bilateral PHC and right TO/LO selectivity for respectively big objects (over body size) and small objects (hand-size object) (see also Bainbridge & Oliva, 2015; Konkle and Caramazza, 2013). Content-dependent fusion enables us to characterize the temporal dynamics of these regions.

Interestingly, whereas the animacy representation lasted during the first 500ms, the object-size representation was less sustained and could be read-out only during the first ~340ms after the stimulus onset.

Conclusion

Content-dependent fusion is a powerful approach to capture the spatiotemporal dynamics of specific cognitive dimensions in the human brain. It extends the MEG-fMRI

Fig 3. Spatiotemporal dynamics of object-size. Content-dependent fusion analyses revealed neural dynamics of object-size across the whole brain over time. For reference, six regions of interest are shown in the top row, obtained from Wang et al. (2015) probabilistic atlas. Red voxels indicate statistical significance (N=15, one-sided sign-rank test, FDR-corrected at 0.05 across time [-100,1101 ms] and space [all voxels]).

fusion approach introduced in Cichy et al (2016), which precisely tracked human brain activity in space and time, by also pinpointing the neural content relevant to cognitive processes represented by a reference model RDM. Here we demonstrated the method in two cases, characterizing the dynamics of animacy and object-size. We revealed the locus and dynamics of both cognitive processes, offering a comprehensive view of brain activation and extending prior MEG and fMRI studies which characterized the neural spatial and temporal dynamics in isolation. By mapping the neural architecture of content information in space and time, the fusion approach presented here shall reveal the nature of specific informational pathways allowing for a glimpse at the bigger picture of how the brain performs the computations necessary for creating visual representations, and inspire computational models of vision

References

Bainbridge, W.A. and Oliva, A., 2015. Interaction envelope: local spatial representations of objects at all scales in scene-selective regions.*NeuroImage*, *122*, pp.408-416.

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

Cichy, R.M., Pantazis, D. and Oliva, A., 2014. Resolving human object recognition in space and time. Nature neuroscience.

Cichy, R.M., Pantazis, D. and Oliva, A., 2016. Similarity-based fusion of MEG and fMRI reveals spatio-temporal dynamics in human cortex during visual object recognition. *Cerebral Cortex*.

Hutchison, R.M., et al., 2013. Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, *80*, pp.360-378.

Khaligh-Razavi, S.M. and Kriegeskorte, N., 2014. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput Biol*, *10*(11), p.e1003915.

Kiani, R., et al., 2007. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. Journal of neurophysiology, 97(6), pp.4296-4309.

Kriegeskorte, N. and Kievit, R.A., 2013. Representational geometry: integrating cognition, computation, and the brain. *Trends in cognitive sciences*, *17*(8), pp.401-412.

Kriegeskorte, N., et al., 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*,*60*(6), pp.1126-1141.

Konkle, and Oliva, A., 2012. A real-world size organization of object responses in occipitotemporal cortex. Neuron,pp.1114-1124.

Konkle, T. and Caramazza, A., 2013. Tripartite organization of the ventral stream by animacy and object size. The Journal of Neuroscience, 33(25), pp.10235-10242.

Naselaris, T., Stansbury, D.E. and Gallant, J.L., 2012. Cortical representation of animate and inanimate objects in complex natural scenes. *Journal of Physiology-Paris*, *106*(5), pp.239-249.

Nili, H., et al. 2014. A toolbox for representational similarity analysis. PLoS Comput Biol, 10(4), p.e1003553.

Yamins, D.L. and DiCarlo, J.J., 2016. Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, *19*(3), pp.356-365.

Wang, L., et al. 2015. Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, *25*(10), pp.3911-3931.

Acknowledgments: This work was funded by National Eye Institute grant EY020484 (to A.O.); NSF award 1532591 and the McGovern Institute Neurotechnology Program (to A.O and D.P) and was conducted at the Athinoula A. Martinos Imaging Center, MIBR, MIT.