Measuring the Spatial Scale of Brain Representations

Avital Hahamy (a.hahamy@ucl.ac.uk)
Wellcome Trust Centre for Neuroimaging, Queen Square 12
London, WC1N 3BG United Kingdom

Tim Behrens (behrens@fmrib.ox.ac.uk)
Wellcome Centre for Integrative Neuroimaging, University of Oxford,
John Radcliffe Hospital, Oxford OX3 9DU, UK

Abstract:

Understanding how the brain encodes information is one of the core questions in cognitive neuroscience. This question has been tackled by measuring fine-grained fMRI activity patterns across voxels, termed brain representations. These measured representations likely capture gross variations in activity across functional sub-regions, which are reflected in patterns of low spatial frequency. However, it is unclear whether patterns that are not driven by functional/anatomical structure (and are therefore expected to contain higher spatial frequencies) also contribute to these representations. Such rugged patterns have the potential to reflect more intricate stimulus-related information. Here we present a novel method for separating the high- from the low-frequency patterns, and evaluating whether these patterns contain reliable information. By relying on cross-subject temporal synchronization of brain activity and within-subject consistency of activity patterns, our method provides evidence that, at least in sensory brain regions, high-frequency patterns hold reliable information. Using the same method we also demonstrate that many of these activity patterns are unique to each individual. These results demonstrate the potential of our novel method to shed new light on the types of information conveyed by brain representation.

Keywords: brain representations; fMRI; idiosyncrasy; spatial frequency

Introduction

Methods of Multi Voxel Pattern Analysis (MVPA, Cox and Savoy, 2003) measure fine-grained fMRI patterns of activity across voxels, termed brain representations. Much research has aimed to infer the information encoded in these patterns, by presenting participants with discrete and well-controlled stimuli (e.g. Mitchell et al., 2004), or by exposing participants to more naturalistic conditions, e.g. naturalistic movies (e.g. Chen et al., 2017). However, the spatial scale of these measured representations is unclear. It is possible that these measured brain representations within a given Region Of Interest (ROI) simply capture gross variations in activity across functional sub-regions contained within that ROI. If so, while such patterns are informative, they may not reflect more fine-grained stimulus-related information that regional brain representations are expected to hold. This problem naturally exists when using "searchlight" analyses, which sample ROIs across the brain, regardless of functional boundaries between regions. Yet, the same problem is also inherent to any definition of ROI, since this definition necessarily relies on arbitrary criteria (activation threshold, ROI size), which are agnostic of the true (and unknown) underlying functional heterogeneity of neurons. So how can we evaluate if brain representations reflect functional/anatomical structures or more fine-grained information?

We propose that these different spatial scales may be distinguishable. Patterns that reflect regional variations in activity are expected to be of a smooth (low frequency) spatial structure, since they are driven by the underlying anatomical/functional structure. Similarly, rugged (higher frequency) patterns of activity are not likely to be driven by anatomy. Yet can such rugged patterns contribute reliable information to the measured representations? These questions form the bases of a novel method that will be presented here. This method allows separating the smooth from the rugged spatial components of brain activity, and evaluating whether these components hold information that may contribute to the formation of brain representations.

Methods & Results

Here we make use of naturalistic stimuli, which evoke rich representations. We made use of freely available fMRI data of n=15 participants who watched a full-length movie during several scanning runs (http://studyforrest.org/). Two randomly chosen runs were analyzed here. Preprocessing of these data were described in Ben-Yakov and Henson (2018). All functional data were masked by two sensory ROIs (left V1 – 162 voxels; left MT/V5 – 112 voxels) and one associative ROI (left Inferior Parietal Lobule, IPL - 104 voxels), all based on a visual localizer task in an independent set of participants (data taken from Wilf et al., 2017).

We first aimed to evaluate the spatial smoothness of representations. To this end, for each ROI, we used

SVD to decompose the fMRI data into 10 spatial components. Visual inspection of these data revealed that early components (associated with high eigenvalues) tend to show smoother representations compared to late components (associated with lower eigenvalues, Figure 1A). To verify this, each spatial component of each participant's data was autocorrelated. Smooth components should yield a relatively wide spread of high autocorrelation values around the autocorrelation peak, whereas autocorrelation of rugged spatial components is expected to result in a relatively narrow spread around the peak (in the extreme case, the only high value will be at the peak itself). As exemplified in Figure 1B, the computed autocorrelations were indeed wider around the peak in early compared to late components, hinting that the level of spatial smoothness decreases across components. To quantify this, we measured the number of values larger than r=0.5 in three planes that pass through the peak (r=1) coordinate of the 3D autocorrelation matrix. The resulting 3 values (one for each main tensor) were averaged to construct an index of spatial "smoothness" for each component and participant. A gradual decrease in the smoothness of early to late components was revealed within each participant using Spearman's correlations, and tested across participants using one-sample t-tests on the fisher-transformed coefficients in each ROI (V1: $\bar{r} = -.73, \mathrm{t}(14) = 13.96$; MT: $\bar{r} = -.67, \mathrm{t}(14) = 11.98$; IPL: $\bar{r} = -.72, \mathrm{t}(14) = 12.5$; all p<.001, Figure 1C).

Having established that later spatial components of each ROI contain a relatively rugged structure, we turned to evaluate whether these later components hold information rather than rugged noise. To evaluate this we developed a novel method, schematically illustrated in Figure 2. This method is based on two criteria that should be met by data produced from movie-viewing fMRI paradigms. *Criterion 1 - Cross-participant temporal consistency:* Two participants (i, j) who watch the same movie-segment (α) should have brain activity that is time-locked to the movie, therefore producing inter-subject temporal correlations of brain activity (Hasson et al., 2004). *Criterion 2 - Within-participant pattern consistency:* a single participant (i) who watches two segments (α , β) of the same movie (segments contain similar stimuli), should have brain representations that are similar across the two viewings. Based on these two criteria, the below described method will aim to predict held out data of participant j watching movie-segment β, by estimating the participant's representations and time-courses evoked by movie β. This method will be applied on patterns of different spatial smoothness, and those that would allow prediction of the held out data (by fulfilling the criteria of inter- and intra-participants consistency) will be considered reliable brain representations.

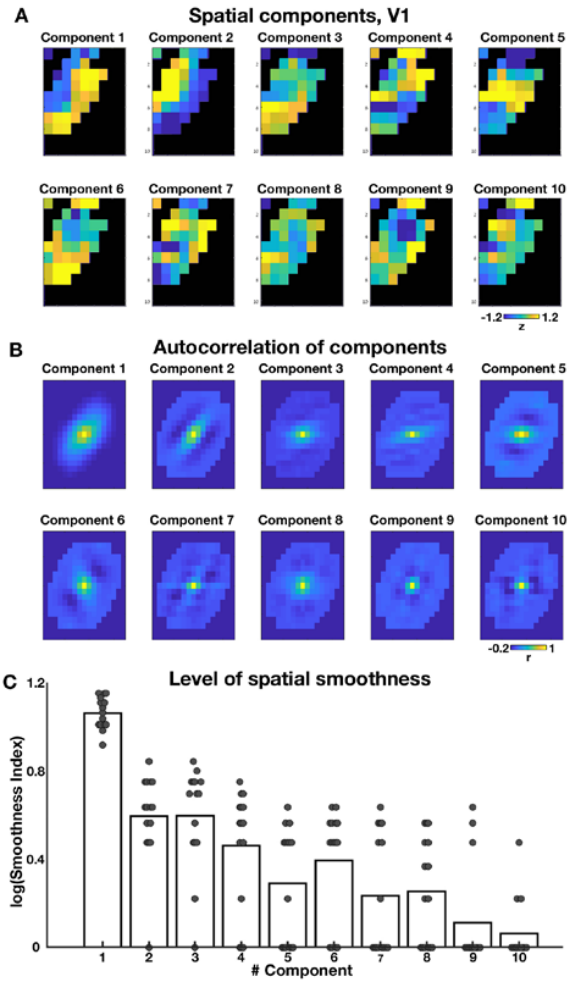

Figure 1. (A) Spatial components of one participant, depicted on a horizontal slice that passes through the center of each component matrix. (B) Autocorrelation of the components presented in (A), depicted on a horizontal slice that passes through the peak of the autocorrelation matrix. (C) Bars depict group-averages, and dots depict the dispersion, of participants' smoothness indices (log scale) within each component. Values of 0 reflect very rugged patterns in which only one voxel (the peak) exceeds r=0.5. Note that correlations were computed within participants.

To obtain brain representations of different smoothness we applied an SVD decomposition to the functional data (voxels*TRs) of participant i watching movie segment α: $\mathrm{D_{i\alpha}} = \mathrm{U_{i\alpha}S_{i\alpha}V_{i\alpha}}^{\mathrm{T}}$, where $\mathrm{D_{i\alpha}}$ denotes the dataset, $\mathrm{U_{i\alpha}}$ denotes the first 10 spatial components and $\mathrm{V_{i\alpha}}$ denotes the first 10 temporal components. The below described method is performed separately on each of the component of $\mathrm{D_{i\alpha}}$.

Based on criterion 1 of *inter-subject temporal consistency*, the time-courses of one participant give an estimation of the time-courses of another participant watching the same movie. We can therefore assume that datasets $D_{i\alpha}$, $D_{j\alpha}$ have common temporal components, $V_{i\alpha}$. If this assumption holds, we could estimate the spatial components of participant j watching movie α, by multiplying data $D_{j\alpha}$ by any of the shared temporal components: $\widehat{U_{j\alpha}} = D_{j\alpha}V_{i\alpha}$.

Based on criterion 2 of *intra-subject pattern consistency*, the brain representations of a certain participant viewing one movie-segment give an estimation of the representations of the same participant while viewing a second movie-segment. We can thus assume that datasets $D_{i\alpha}$, $D_{i\beta}$, have shared stimulus representations (spatial components), $U_{i\alpha}$. If this assumption holds, by multiplying the data $D_{i\beta}$ by any of these spatial components we could estimate the temporal components of participant i watching movie β: $\widehat{V_{i\beta}} = {D_{i\beta}}^{T}U_{i\alpha}$.

This procedure results in an estimation of a specific temporal component of movie β ($\widehat{V_{i\beta}}$) that is shared between participants watching the same movie β (Hasson et al., 2004), as well as an estimation of a specific spatial component of subject j ($\widehat{U_{j\alpha}}$) that is shared across data acquired from that same participant viewing different movie segments. We can therefore estimate the data of participant j watching movie β, by multiplying the estimated temporal and spatial components that compose this dataset: $\widehat{D_{j\beta}} = \widehat{U_{j\alpha}}{\widehat{V_{i\beta}}}^{T}$. Critically, we have access to the actual held out data $D_{j\beta}$. We can therefore assess the quality of our prediction by correlating the estimated data, $\widehat{D_{j\beta}}$, with the real data, $D_{j\beta}$: $\rho_{\widehat{D_{j\beta}},D_{j\beta}}$.

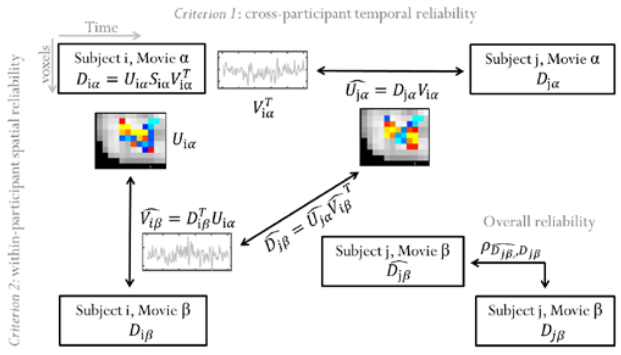

Figure 2. Schematic description of the analytic pipeline. The horizontal branch is based on Criterion 1, the vertical branch is based on Criterion 2.

The resulting component-specific correlation coefficients provide an estimation of the level to which each component holds reliable inter- and intra-participant information, whose combination allows to predict a new dataset. As detailed in Table 1, many of the rugged, as well as smooth, components of the sensory ROIs proved to be significantly reliable, as measured using one-sample t-tests across the Fisher transformed correlations of all subject-pairs. Reliability of some rugged IPL components showed only trend-level significance, as they did not survive Bonferroni correction for 30 comparisons.

Table 1: Mean correlation coefficients (estimating reliability) and uncorrected p-values per components (rows) and ROIs (columns). Note that asterisked p-values mark tests that survived Bonferroni correction.

#Comp/ROI	V1	MT	IPL
1	.21 (<.001*)	.52 (<.001*)	.12 (<.001*)
2	.08 (<.001*)	.09 (<.001*)	.03 (.001*)
3	.07 (<.001*)	.1 (<.001*)	.03 (<.001*)
4	.04 (<.001*)	.04 (<.001*)	.02 (<.001*)
5	.04 (<.001*)	.07 (<.001*)	.02 (.03)
6	.04 (<.001*)	.03 (.009)	-0.0007 (.92)
7	.03 (<.001*)	.05 (<.001*)	-0.002 (.76)
8	.04 (<.001*)	.03 (<.001*)	.01 (.06)
9	.01 (.11)	.08 (<.001*)	.02 (.008)
10	-0.01 (.33)	.06 (<.001*)	.02 (.02)

Finally, our method allows to study whether brain representations are idiosyncratic (individually-unique) or shared across participants (Chen et al., 2017). Specifically, if brain representations are shared across participants, then using a dataset of any other participant watching the movie segment β, $D_{k\beta}$, instead of using $D_{i\beta}$ in the analytic pipeline, will result in just as good of an estimation of the temporal components $V_{i\beta}$, and therefore in a similar correlation between $\widehat{D_{j\beta}}$, and $D_{j\beta}$. We used this assumption of shared representations as the null hypothesis in a permutation test, which allowed us to shuffle the identities of participants in the selection of $D_{i\beta}$. Rejection of this null hypothesis would suggest that a certain spatial component holds individually unique information.

To this end, for each component, the same analytic pipeline was applied using all possible pairs of

participants. This resulted in a symmetrical matrix (n*n) of correlation coefficients. The test statistic was defined as the mean of the lower triangular part of this matrix. Next, 10,000 random correlation matrices were generated by randomly drawing $D_{k\beta}$ in the analytic pipeline calculated for each pair of participants i,j. The mean of the lower triangular part of each random matrix was calculated, and these 10,000 random means constructed the null distribution. The position of the test statistic in relation to this null distribution was used to derive a two-sided p-value. All p-values (for components and ROIs) were Bonferroni corrected for 30 comparisons. As detailed in Table 2, many of the spatial components of the sensory ROIs were found to hold idiosyncratic information. Some effects were also detected in the IPL components, though most did not survive Bonferroni correction for multiple comparisons.

Table 2. Idiosyncrasy uncorrected p-values per components (rows) and ROIs (columns). Asterisks mark tests that survived Bonferroni correction.

#Comp/ROI	V1	MT	IPL
1	0.45	<.001*	.37
2	<.001*	<.001*	.22
3	0.4	<.001*	<.001*
4	0.33	0.31	.22
5	<.001*	<.001*	.02
6	0.009	.001*	.27
7	<.001*	<.001*	.3
8	<.001*	<.001*	.04
9	0.07	<.001*	.1
10	0.33	0.009	.13

Discussion

Here we demonstrated that brain representations are composed of both high and low spatial frequencies. We assumed that while the smooth spatial components represent anatomically-related activity variations, the rugged components may hold intricate stimulus-related information. Using a novel method, based on measures of spatiotemporal inter- and intra-participant reliability, we demonstrated that many of the rugged (and smooth) components hold reliable information. We further reveal that many components hold idiosyncratic information.

It is important to note that we make no claim about the specific spatial components or the particular functional information they may encode, as the sorting of components by their eigenvalues is not necessarily identical across participants. Nevertheless, we demonstrate that the serial position of components is related to their level of smoothness, and therefore propose that reliable information can be encoded not only in the smooth, but also in the rugged components.

The significance of our reported effects depended on the specific brain regions examined. Indeed, rugged IPL components showed far weaker idiosyncrasy effects compared to the components of visual regions. These results are in agreement with previous work, suggesting that areas of the Default Mode Network (like the IPL) have shared representations across participants who watch the same movie (Chen et al., 2017). While these previous findings may have been driven by smooth spatial components that reflect shared functional/anatomical structure, the trend-level reliability effects of the rugged IPL components in our data (Table 1) imply that rugged IPL patterns may still play a functional role (regardless of their idiosyncrasy) in the encoding of information. Further research is therefore needed for characterizing the nature of representations in such associative brain areas.

Acknowledgements

AH was supported by a European Molecular Biology Organization non-stipendiary Long-Term Fellowship (848-2017), Human Frontier Science Program fellowship (LT000444/2018), and a Marie Curie Individual Fellowship (789040). TB was supported by a Wellcome Trust grant (HMR00560.001).

References

Ben-Yakov A, Henson RN (2018) The Hippocampal Film Editor: Sensitivity and Specificity to Event Boundaries in Continuous Experience. J Neurosci 38:10057-10068.

Chen J, Leong YC, Honey CJ, Yong CH, Norman KA, Hasson U (2017) Shared memories reveal shared structure in neural activity across individuals. Nat Neurosci 20:115-125.

Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19:261-270.

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. Science 303:1634-1640.

Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang X, Just M, Newman S (2004) Learning to decode cognitive states from brain images. Machine learning 57:145-175.

Wilf M, Strappini F, Golan T, Hahamy A, Harel M, Malach R (2017) Spontaneously Emerging Patterns in Human Visual Cortex Reflect Responses to Naturalistic Sensory Stimuli. Cereb Cortex 27:750-763.