Supplementary Information for:

Cortical Networks of Dynamic Scene Category Representation in the Human Brain

Emin Çelik[a,b], Umit Keles[b,c], İbrahim Kiremitçi[a,b], Jack Gallant[d,e,f], Tolga Çukur[a,b,g]

[a]Neuroscience Program, Sabuncu Brain Research Center, Bilkent University, Ankara, TR-06800, Turkey

[b]National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, TR-06800, Turkey

[c]Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA-91125, USA

[d]Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720, USA

[e]Program in Bioengineering, University of California, Berkeley, CA 94720, USA

[f]Department of Psychology, University of California, Berkeley, CA 94720, USA

[g]Department of Electrical and Electronics Engineering, Bilkent University, Ankara, TR-06800, Turkey

Table of Contents

Supplementary Text

1. Potential Confounding Factors

Supplementary Figures

1. Supp. Figure 1: Comparison of the scene-category and part-of-scene models.
2. Supp. Figure 2: Variance in voxelwise tuning profiles explained by voxel cluster centers.

Supplementary Tables

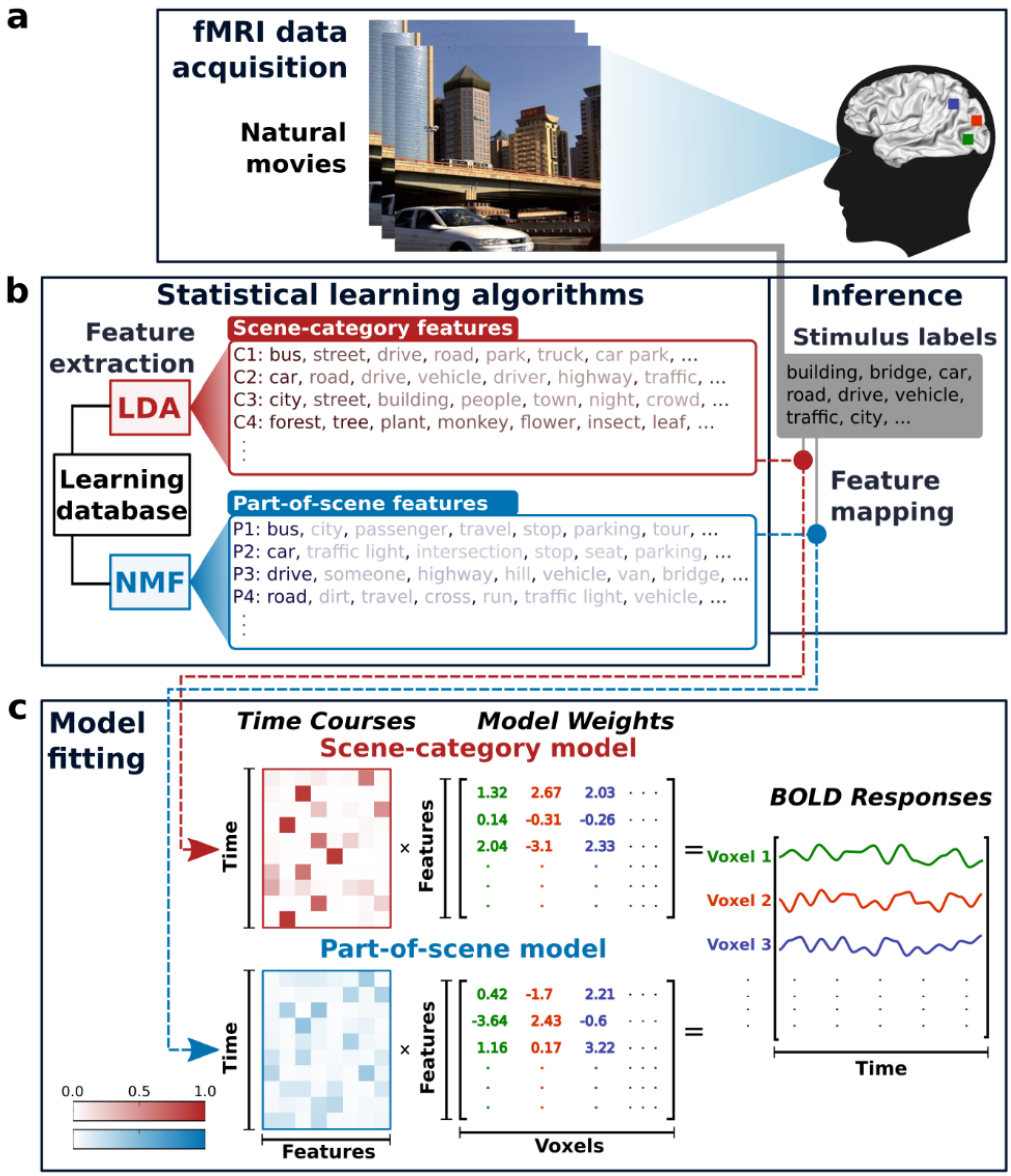
1. Supp. Table 1: Proportion of voxels where the scene-category model outperforms the control model and vice versa.

Supplementary Text

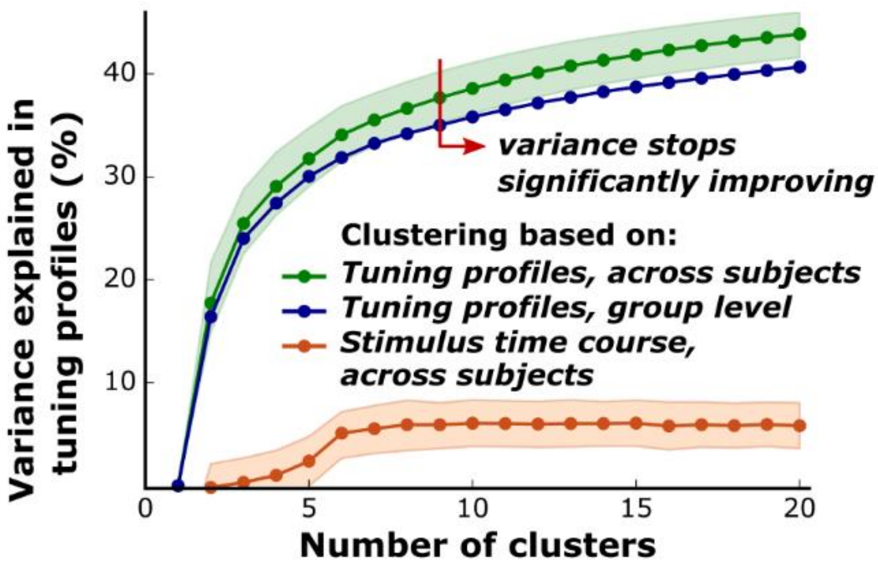
Potential Confounding Factors

In complex natural scenes, some scene categories might co-occur. For instance, a scene depicting "a group of climbers that are talking on a mountainside" could be interpreted as "a mountain scene", "a social communication scene", and "a sports activity scene". As the natural movies used here were of finite length, it is likely that there is a stimulus sampling bias for scene categories that tend to co-occur. If this sampling bias is large, the estimated tuning profiles and subsequent cluster analyses might be biased as well. To rule out this potential confound, we performed a control analysis. To characterize the sampling bias, we performed cluster analysis on the stimulus time course of scene-category features, and obtained stimulus cluster centers (see Materials and Methods). We reasoned that if co-occurrence of scene categories in the stimulus does not induce a significant sampling bias in scene-category tuning profiles, then the cluster centers based on tuning profiles should explain greater variance in tuning profiles than the stimulus cluster centers. We find that in all subjects the cluster centers based on tuning profiles explain significantly higher variance than stimulus cluster centers (bootstrap test, $p < 0.05$; Supp. Figure 2). This result indicates that the nine voxel clusters and their estimated scene-category tuning profiles do not merely reflect sampling bias for scene categories in the movies.

Supplementary Figures



Supp. Figure 1. Comparison of the scene-category and part-of-scene models. a, Whole-brain BOLD responses were recorded while subjects passively viewed two hours of natural movies. **b,** Two separate models were fit to individual voxels to assess scene category representations of natural movies across neocortex. A scene-category model was used to assess representation of scene categories. A part-of-scene model was used to assess representation of constituent object and action components of natural scenes. Model features were extracted via unsupervised learning on a large corpus of natural scene annotations. Scene-category features (C1, C2 etc.) were extracted using LDA in order to capture co-occurrence statistics of objects and actions in dynamic natural scenes. Part-of-scene features (P1, P2 etc.) were extracted using NMF in order to capture constituent groups of objects and actions in natural scenes. Each model feature is defined as a list of probabilities that reflect the likelihood of individual objects and actions occurring in a scene. (Font weights for object-action categories reflect their respective probabilities.) **c,** Salient objects and actions in each 1-s clip of the movies were labeled manually. The movies were then projected separately onto scene-category and part-of-scene features to determine stimulus time courses of the respective models. Regularized linear regression was used to fit voxelwise models that optimally predict BOLD responses in individual voxels. The estimated model weights characterize the tuning of individual voxels for distinct model features.



Supp. Figure 2. Variance in voxelwise tuning profiles explained by voxel cluster centers. To assess the cortical organization of scene category representation, cortical voxels in individual subjects were clustered according to their scene-category tuning profiles. The proportion of variance in tuning profiles explained by cluster centers was examined to determine the optimal number of clusters. For each number of clusters, the proportion of explained variance was first computed in individual subjects and then averaged across subjects (green line, mean ± sem across subjects). To increase sensitivity, the proportion of explained variance was also computed at the group level by pooling voxels across subjects (blue line). The optimal number of voxel clusters was determined as nine because beyond that point, the improvement in explained variance due to an additional cluster fell below 1% in both individual-subject and group-level analyses. To ensure that the clusters based on tuning profiles do not merely reflect a sampling bias for scene categories in the movies, the proportion of variance explained in tuning profiles by stimulus cluster centers (orange line, mean ± sem across subjects) was compared to that explained by voxel cluster centers (see Materials and Methods). Voxel cluster centers explain a much greater proportion of variance compared to the stimulus cluster centers ($p < 0.05$, bootstrap test). This result suggests that the identified voxel clusters are not unduly affected by stimulus sampling bias.

Supplementary Tables

% voxels	S1		S2		S3		S4		S5		Aggregate	
	SC	CM	SC	CM	SC	CM	SC	CM	SC	CM	SC	CM
RET	53.68	26.91	27.26	55.09	24.76	55.47	33.43	48.20	74.44	12.26	42.71±9.42	39.59±8.59
V7	56.96	19.13	5.71	81.71	16.48	66.48	63.35	19.92	39.06	40.63	36.31±11.17	45.57±12.50
LO	50.00	6.25	5.26	86.84	3.45	86.21	87.93	5.17	65.00	15.00	42.33±16.64	39.89±19.11
MT+	49.42	13.87	8.54	75.38	22.92	60.42	56.36	24.55	48.97	18.52	37.24±9.17	38.55±12.33
EBA	53.38	14.29	2.40	74.40	11.84	57.89	64.81	8.58	45.06	18.39	35.50±12.09	34.71±13.19
FFA	16.13	29.03	7.32	58.54	6.25	59.38	52.17	6.52	30.77	30.77	22.53±8.61	36.85±10.00
PPA	88.46	3.85	62.71	10.17	72.92	10.42	90.00	0	74.56	7.89	77.73±5.12	6.47±2.00
OPA	79.41	2.94	6.00	56.00	46.15	23.08	100.00	0	90.74	0	64.46±17.22	16.40±10.80
RSC	65.45	18.18	91.30	0	82.43	8.11	79.17	4.17	68.57	11.43	77.38±4.70	8.38±3.11
IPS	61.33	21.90	18.53	63.53	40.43	34.04	76.67	9.39	69.91	11.11	53.37±10.63	28.00±9.92
pSTS	25.34	28.05	24.13	41.90	33.45	34.89	63.83	10.64	47.52	16.34	38.85±7.51	26.36±5.76

Supp. Table 1. Proportion of voxels where the scene-category model outperforms the control model and vice versa. Percentage of voxels where the scene-category model outperforms the control model (SC) and vice versa (CM) among voxels significantly predicted by either of the two models. Results are given for eleven ROIs in all subjects ($\Delta r > 0.02$). Aggregate values are also reported as mean ± SEM across five subjects.