An open dataset of functional MRI responses to egocentric navigation through natural scenes

Emily M. Chen* (echen15@mit.edu), Frederik S. Kamps* (fkamps@mit.edu), and Rebecca R. Saxe (saxe@mit.edu)

Department of Brain and Cognitive Sciences, 43 Vassar Street, Building 46-4021
Cambridge, MA 02139, United States of America
* *These authors contributed equally.*

Abstract

Successful visually-guided navigation is crucial for everyday life, yet how the human visual cortex responds during dynamic, egocentric navigation is understudied. Here, we created a dataset of functional MRI responses to the egocentric visual experience of visually-guided navigation. This dataset includes voxelwise fMRI responses estimated for 172 unique dynamic scene videos and 18 control videos of faces, objects, or abstract patterns. Scene videos depicted unfamiliar places with no visible people and varied in ego-motion direction, affordances, and scene content. Participants also completed a separate set of functional localizer scans. Univariate analyses of subject-specific scene-selective regions revealed clear scene selectivity at the individual stimulus level, as well as considerable reliable variation in the strength of response across different scene videos. The top-performing videos, which could be used to optimize the efficiency of localizer tasks, tended to depict dynamic ego-motion through doorways and tight spaces with numerous viewpoints, whereas bottom-performing videos tended to depict limited ego-motion through open spaces with little change in viewpoint. Next, multivariate analyses showed evidence of region-specific scene representations, consistent with previous work finding functional distinctions between scene regions. These features set the stage for future experiments using this dataset to test representations of naturalistic variation in navigational information, as well as comparing computational models to human brains.

Keywords: Navigation, vision, fMRI, natural images, scene perception, occipital place area, parahippocampal place area

Introduction

The human visual system supports visually-guided navigation with remarkable accuracy and flexibility, still outperforming state-of-the-art artificial systems in robotics and computer vision. How does the human visual system achieve this feat? As an initial answer to this question, several decades of work in cognitive neuroscience have revealed a network of at least three cortical regions dedicated to representing visual scene information, including the parahippocampal (PPA), medial (MPA), and occipital (OPA) place areas (Epstein & Baker, 2019). All three regions respond selectively to visually presented scenes compared with other visual categories (e.g., faces or objects). However, these regions also show functional dissociations. For example the OPA is significantly more sensitive to dynamic scene information than PPA or MPA (Kamps, Lall, & Dilks, 2016) and is hypothesized to play a critical role in visually-guided navigation, whereas the PPA is hypothesized to support scene categorization and the MPA memory-guided navigation (Dilks, Kamps, & Persichetti, 2022). Despite the fact that we regularly experience our visual world dynamically during navigation, almost all studies to date have measured responses in these regions to static images. Thus, how these regions respond during the dynamic experience of egocentric visually-guided navigation remains unclear.

To facilitate the study of dynamic scene processing, we collected a dataset of responses to 190 unique video stimuli. The majority of stimuli (N=172) depicted the egocentric visual experience of moving through a naturalistic scene. Scenes varied in egocentric motion direction (e.g., straight ahead, turn left or right), affordances (e.g., path or doorways to left or right), openness (e.g., field vs corridor), and content (e.g., forest, city, house, classroom).

Results

Methods Five subjects (M_{age} = 28 years, two female, three right-handed) were recruited from the Greater Boston Area via convenience sampling. fMRI data were acquired from a 3-Tesla Siemens Magnetom Prisma scanner located Athinoula A. Martinos Imaging Center at MIT, using a 32-channel head coil. Across two scan sessions scheduled within a week of one another, participants viewed 20 runs of stimuli presented in an event-related design and consisting of 190 3-second videos, with 172 videos of scenes and 18 control videos (9 object, 3 face, and 6 baseline). Across the two sessions, participants viewed 8 repetitions per video. Participants also completed 4 runs of a functional localizer experiment in which a separate set of scene, face, object, and scrambled object videos were presented in a block design. Using data from the localizer experiment, scene-selective functional regions of interest (fROIs) were defined for each individual subject's brain based on the contrast of scenes$>$objects, face areas based on the contrast of faces$>$objects, and object areas based on the contrast of objects$>$scrambled objects. Early visual cortex (EVC) was defined anatomically based on probabilistic parcels (Wang, Mruczek, Arcaro, & Kastner, 2015). Data were preprocessed using fMRIprep (Esteban et al., 2019), and betas for individual video stimuli were estimated using GLMsingle (Prince et al., 2022).

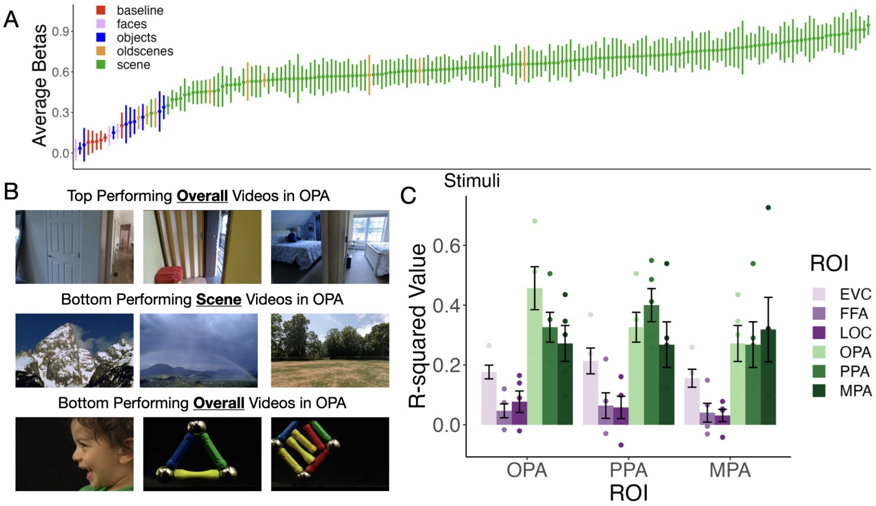

Figure 1: **(A)** Average responses across five subjects for each of the 190 stimulus videos. Notably, a majority of the scene videos created specifically for this experiment drive higher responses relative to scene videos previously tested in other work. **(B)** Example frames from the videos that elicited the overall highest responses (top row), the lowest responses amongst the scene videos (middle row), and the overall lowest responses (bottom row) in OPA. **(C)** Results from a multivariate analysis showing the average R-squared values across five subjects between scene regions and itself compared to other scene and control regions.

Univariate Analyses Univariate analyses revealed clear scene selectivity at the individual stimulus level (Figure 1A). The majority of the scene videos elicited numerically stronger responses than the top-performing non-scene stimulus in all three scene regions (e.g., 168/172 of the scene videos elicited a higher response in OPA than the top-performing non-scene video). Univariate responses were highly reliable, with inter-subject correlations $>$ 0.7 for all three scene regions (measured as the Pearson correlation between each subject's univariate responses across videos and the average of the remaining 4 subjects' univariate responses). Responses also varied considerably across the different scene videos, with the top-performing scene videos eliciting 1.7-2.5 times stronger responses than the bottom-performing scene videos across scene regions. Videos eliciting the strongest activation depict navigation through tight boundaries and doorways between multiple spaces, whereas videos eliciting the weakest activation depict static views or stable forward motion through open spaces with distant boundaries or no visible ground plane (Figure 1B). These patterns of results were similar in all three scene regions, with no ROI-specific information detected; the correlation of univariate responses in a scene region to itself (measured across split halves of the dataset) was not stronger than itself with other scene regions.

Multivariate Analyses Multivariate analyses were performed by computing a representational similarity matrix in each fROI in each subject, with cells comprised of the Pearson correlation (across voxels) for all possible pairs of videos. For each subject, we assessed the similarity of these representational spaces between regions by computing the correlation of the bottom triangle of each matrix between each region and each other region (including itself) across split halves of the data (Figure 1C). This analysis revealed that scene regions are more strongly correlated with other scene regions than face, object, or early visual regions, but additionally that OPA and PPA are more strongly correlated with themselves than with the other scene regions ($\beta = 0.158, SE = 0.044, p = 0.0071$ for OPA compared to PPA and MPA and $\beta = 0.103, SE = 0.0425, p = 0.0418$ for PPA compared to OPA and MPA). This pattern of results held when all video stimuli were included and also when the analyses was limited to scene videos only. These results suggest that this dataset captures both shared and unique aspects of the representational spaces encoded in each region of the scene network, consistent with previous work suggesting that scene regions are all scene-selective but nevertheless play distinct roles in scene processing.

Discussion

We present a novel, condition-rich dataset of fMRI responses to short videos depicting the egocentric experience of visually-guided navigation through unfamiliar scenes, with variability in ego-motion and scene content. Whole-brain, voxelwise response estimates for 190 individual videos across 8 repetitions per video in 5 subjects will be made publicly available upon project completion. Univariate and multivariate analyses confirm that this dataset is well-suited for experiments exploring how naturalistic scene information experienced during navigation is represented across the cortical scene network and for comparisons between human brain responses and computational models. Furthermore, we identify a subset of top-responding videos that can be used for an efficient and engaging localizer experiment, with promising use cases for diverse (e.g., pediatric, neuropsychological) populations.

References

Dilks, D. D., Kamps, F. S., & Persichetti, A. S. (2022). Three cortical scene systems and their development. *Trends in Cognitive Science*, *26*, 117–127.

Epstein, R. A., & Baker, C. I. (2019). Scene perception in the human brain. *Annual review of vision science*, *5*, 373–397.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., . . . Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*, 111–116.

Kamps, F. S., Lall, V., & Dilks, D. D. (2016). The occipital place area represents first-person perspective motion information through scenes. *NeuroImage*, *83*, 17–26.

Prince, J. S., Charest, I., Kurzawski, J. W., Pyles, J. A., Tarr, M. J., & Kay, K. N. (2022). Improving the accuracy of single-trial fMRI response estimates using GLMsingle. *eLife*, *11*, e77599.

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, *25*, 3911–3931.