Arrow of Time:
an intensive fMRI dataset of whole-brain responses to brief videos

Anonymous authors
Double blind review

Abstract

Extensive sampling of neural activity during rich cognitive phenomena is critical for robust understanding of brain function. To this end, we introduce the Arrow of Time dataset, a novel resource that captures high-quality functional magnetic resonance imaging (fMRI) responses at 7T to thousands of richly annotated short (2.5s) naturalistic videos. Our goal is to provide a deep open-by-design dataset of brain responses to the field.

Participants are exposed to up to 2000 video stimuli while ensuring strict fixation, across up to 15 scanning sessions. The extensive and varied stimuli are annotated using neural network-based object segmentation, action recognition, and semantic descriptions. This allows researchers to investigate how the brain processes and represents complex visual information, such as object recognition, scene understanding, and semantic processing. Initial quality assessments show noise ceiling values on a par with those in NSD.

This type of high-quality fMRI data allows for rigorous testing and refinement of models related to brain function and cognition. As a reusable resource, the Arrow of Time dataset will help elucidate the neural basis of naturalistic visual processing in the hands of researchers in the fields of cognitive neuroscience, psychology, and artificial intelligence.

Keywords: fMRI, naturalistic visual processing, video, 7-Tesla

Introduction

A new paradigm for fMRI is to deeply sample an individual's brain responses to rich stimuli across many sessions. These open-by-design "intensive" datasets can be used to bridge between AI and neuroscience, and provide a unique perspective on neural function. One benchmark example of such an intensive dataset is the Natural Scenes Dataset (NSD), which used up to 10000 presentations of still natural images and scanned at 7 Tesla for increased SNR, in 8 subjects. Subsequent datasets have extended this paradigm to video presentation, which provides a more engrossing perceptual experience than still images, and results in broader swathes of activation across the cerebral cortex. Examples are the BOLD moments dataset (10 subjects) used 3 Tesla fMRI and used >1000 brief YouTube videos as stimuli, whereas the stimulus material of the Human Connectome Project 7 Tesla dataset (180 subjects) consists of an hour of ~3 min audiovisual video segments.

Here, we use 7 Tesla fMRI to measure whole-brain responses to up to 2000 brief videos, taking advantage of the increased signal fidelity at higher field strengths while capitalizing on the brevity of the video material to assert experimental control.

Data Acquisition & Analysis

At least 8 subjects will be scanned for up to 15 sessions of 7 Tesla (f)MRI. The main video experiment will take up 12 of these sessions. These sessions will be supplemented by a full session of quantitative anatomical MRI at 7T for function-histology comparisons and a full session at 3T for T1w and T2w anatomies for FreeSurfer reconstruction. The remaining functional sessions will contain: 1. a replication of the HCP movie-watching experiment for

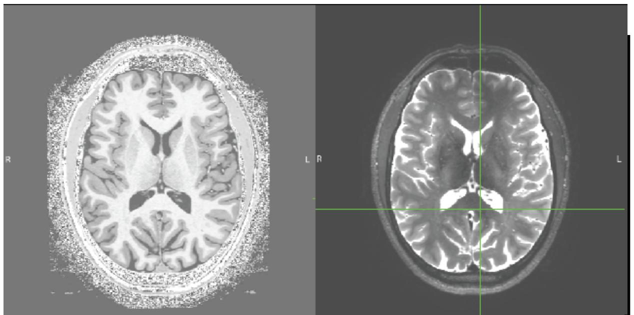

Fig 1: high-resolution 7 Tesla T1w and T2w anatomies

comparison with extant data, 2. A high-powered pRF mapping experiment for delineation of retinotopic maps, 3. Resting State scans.

Each functional session contains a sub-millimeter resolution T2*-weighted anatomical scan for alignment across acquisition sessions, and multiple separate acquisitions for susceptibility distortion correction. All fMRI data acquisition occurs at 1.6s TR and 1.7mm isotropic resolution, with whole-brain coverage. We have created a tailored preprocessing pipeline to estimate distortion fields due to gradient nonlinearity & susceptibility and motion / coregistration parameters, and perform a single interpolation step to the anatomical space for minimal implicit smoothing. BOLD time-courses are temporally up-sampled, and entered into a GLMsingle analysis, which a) estimates single-trial beta weights while b) projecting out noise sources and c) estimating the HRF shape at the single-voxel level. Data quality quantifications show noise ceiling levels on a par with NSD (Fig 2). Once completed, we will provide multiple surface and volume-based formats of the single-trial beta weight estimates for download, with separate versions with and without having undergone NORDIC denoising.

Experiment & Stimulus materials

2200 2.5s videos were sourced from creative commons databases and chosen to portray a sampling of different scenes, human actions and emotion, and event types. Videos were

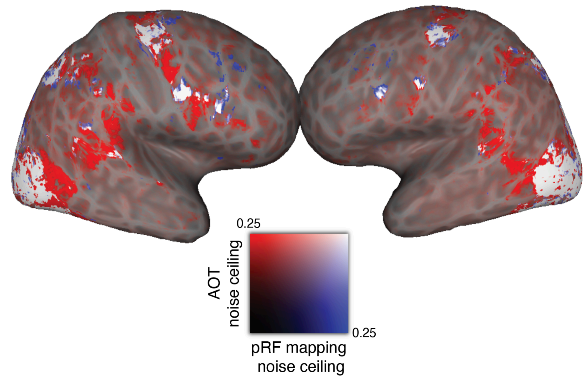

Fig 2: Data quality quantification, comparing AOT noise ceiling with separate session pRF noise ceiling. This indicates that more anterior regions of the visual cortex respond to naturalistic video than do to the bar-shaped stimulus used for retinotopic mapping.

resampled and identically encoded to be shown at 1080p, 24Hz, ensuring equal video quality across all videos. We provide an array of stimulus annotations, such as semantic annotations, motion energy content, depth mapping, bodypart keypoints and more.

2.5s videos were shown separated by 900ms fixation intervals, across 10 runs in each 1-hour scanning session. Participants performed a memory encoding task during scanning, tested in a yes-no recognition experiment after scanning. To ensure fixation during video presentation, we used a gaze-contingent conditioning experiment to train our participants to fixate accurately during presentation of even the most dynamic video materials before the experiment, and recorded gaze using IR eyetracking throughout.

Conclusions

The AOT dataset will provide the field with a deep video-fMRI dataset of unparalleled quality and scope. We consider AOT a pilot for further development of intensive fMRI datasets in the lab, and look forward to discussing ideas for further intensive dataset acquisition with the field.

References

Kupers, E. R., Knapen, T., Merriam, E. P. & Kay, K. N. Principles of intensive human neuroimaging. *Trends Neurosci.*(2024) doi:10.1016/j.tins.2024.09.011.

Allen, E. J. *et al.* A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat Neurosci*1–11 (2021) doi:10.1038/s41593-021-00962-x.

Lahner, B. *et al.* Modeling short visual events through the BOLD moments video fMRI dataset and metadata. *Nat. Commun.* **15**, 6241 (2024).

Dumoulin, S. O. & Wandell, B. A. Population receptive field estimates in human visual cortex. *NeuroImage* **39**, 647 660 (2008).

Prince, J. S. *et al.* Improving the accuracy of single-trial fMRI response estimates using GLMsingle. *Elife* **11**, e77599 (2022).

Vizioli, L. *et al.* Lowering the thermal noise barrier in functional brain mapping with magnetic resonance imaging. *Nat Commun* **12**, 5181 (2021).

Dumoulin, S. O. & Knapen, T. Ultra-High Field Neuro MRI. *Adv. Magn. Reson. Technol. Appl.* **10**, 407–418 (2023).