Title

# Abstract

Abstract text

# Introduction

Decoding cognitive states (refs) from fMRI opens up exciting new applications in training and therapy. However, many hurdles to clear before a real-world application is possible.

Training or therapy programs must be performed within the scanner. Virtual reality allows for broader range of possible programs. Some therapies already utilize virtual reality. See for example PTSD extinction therapy (refs). Cognitive state decoding must be accurate and reliable without imposing too many restrictions on the stimulus.

Most cognitive neuroscience experiments are focused on explanatory power at the expense of classification accuracy and stimuli complexity (makes sense when trying to answer a research question). However, our goal is to maximize decoding accuracy in a complex environment that could be used for training or therapy. Therefore, we propose the use of more complex multi-variate decoding techniques. They are more difficult to interpret but provide improved decoding accuracy.

Unfortunately, fMRI data is relatively ill posed for advanced machine learning methods. The dimensionality of whole brain data is large compared to the number of samples we can collect. Furthermore, the samples are correlated in both space and time. Many data-agnostic approaches have been used to solve this problem such as PCA and ICA. However, we want to use the structure of fMRI data and more specifically the brain to intelligently reduce the dimensionality of the data as much as possible before using something like PCA or ICA. We project the functional data on to extracted cortical surfaces by averaging through the gray matter normal to the surface. Only looking at signal from gray matter reduces the dimensionality of the data by a factor of #. Then we spatially smooth only along the surface thereby avoiding averaging our data with unwanted noise in the white matter and CSF. Furthermore, this also prevents averaging across sulcal/gyral boundaries where we no longer expect the signal to be spatially correlated. Finally, we down-sample on the surface based on our smoothing parameter to reduce the dimensionality again.

Most studies report within-session decoding accuracy. That is, the decoding accuracy on a held out test set from the same scanning session that the training data was collected in. It’s important to be able to train the classifier in one session, and then use that trained classifier in subsequent sessions (and potentially train a classifier on multiple sessions and subjects and apply it to a subsequent session). To solve this problem, we have found that spherical registration with surface based smoothing and down-sampling yields the best results across both sessions and subjects. This cross-session registration procedure also allows us to collect and utilize more training examples for machine learning which also helps with the disparity in dimension and training examples and yields improved classification accuracy.

We want to decode an internal variable, decoding the presentation of a visual or audio stimuli are uninteresting for our purposes; we already know what stimuli we are presenting to the subject. Instead, we propose to decode the subject’s performance. This is something we don’t already know when we present the stimuli (though we can predict to a certain degree from the difficulty setting of the task). Furthermore, predicting a subject’s performance has obvious applications for training and therapy such as modulating difficulty to keep predicted performance on a specific trajectory.

In traditional cognitive neuroscience experiments, fMRI signals are averaged across repeated presentations of the same stimuli to boost signal to noise. The stimuli is highly controlled to ensure that the elicited cognitive state is the same (or nearly the same) across all presentations. However, since we don’t know *a priori* what the elicited internal cognitive state will be (by design; if we already knew the state then why are we decoding at all?) we must develop techniques that can leverage the temporal structure of the signal to deal with the signal to noise problem and improve decoding accuracy. Hidden Markov models have been used in a variety of fields to solve precisely this problem (refs).

# Methods

## Subjects

Six adult males, ages 24-57 (TODO: need to check how old Ethan was), with normal or corrected-to-normal vision, participated in the experiments. All subjects participated in two fMRI sessions and a third session to acquire a high-resolution structural anatomy. Informed consent was obtained from all subjects under a protocol approved by the University of Texas at Austin Institutional Review Board.

## Stimulus

Given our long-term interest in PTSD, we created a virtual town intended to suggest a setting encountered by military forces.

Subjects perform a weapon detection task. 20 characters move through subject’s view. Characters pull out a weapon every 2—4 s. Subjects press a button to indicate they detected the weapon.

Task difficulty controlled by varying the distance of the character with the weapon and the presentation time.

Performance data collected for each subject to tune the difficulty.

Based on the collected psychophysics data, visibility of the weapon was measured each frame to adjust presentation time and keep the difficulty constant

Low-level contrast adjusted in real-time to minimize correlation with task difficulty

Task performance calculated as average performance during block.

Performance values binned into three categories: poor, average, and good.

## Psychophysics

Trained outside of scanner

Sigmoid fit to performance curves

Game tuned to provide maximum difference between easy, medium, and hard difficulties

## MRI Protocols

Imaging was performed on a Sieman’s Skyra 3T scanner using the product 32-channel head coil. Structural reference volumes were T1-weighted with good gray-white contrast and acquired using a 3D inversion-prepared fSPGR sequence (minimum TE and TR, TI = 450 ms, 15deg flip angle, isometric voxel size of 0.7 mm, 2 excitions, ~28-min duration). fMRI scans were collected using a whole-brain GRAPPA EPI sequence with g-factor = 2, TE = 25 ms, TR = 2.5s, and 2-mm cubic voxels across 200mm field-of-view. The slice prescription included 60 slices oriented along the AC-PC axis. A high-order shim was performed before the start of the functional imaging to improve field homogeneity. A set of T1-weighted structural images were obtained on the same prescription as the functional acquisition runs in the same session directly before the functional scans were collected. Images were acquired using a three-dimensional (3D) fast RF-spoiled gradient-echo (fSPGR) sequence. These anatomical images were then used to align the functional data to the structural reference volume.

## Preprocessing

Automatic cortical segmentation and surface extraction was performed on the structural reference volume using FreeSurfer (ref). The cortical surfaces for each hemisphere were inflated into a sphere while minimizing metric distortion. These spherical surfaces are then registered to a spherical atlas first by coarsely aligning on large-scale folding patterns and then fine-tuned using small-scale curvature patterns (freesurfer sphere ref).

The inplane anatomical volumes were skull-stripped and normalized in the same fashion as the first stage of the automatic cortical segmentation and surface extraction. The processed inplane volumes were then affinely registered to the structural reference volumes using a method based on robust statistics to detect outliers and remove them from the registration (Highly Accurate Inverse Consistent Registration: A Robust Approach).

Simultaneous slice-timing and motion correction was performed on the functional scans (ref nipy.SpaceTimeRealign). Then, a rigid-body registration was performed between scans to align each frame to the first volume, i.e., the frame closest in time to the structural inplane (Improved Optimisation for the Robust and Accurate Linear Registration and Motion Correction of Brain Images).

The functional data was then approximately aligned to the structural reference volume using the previously calculated registration with the inplane anatomical data. A boundary based registration technique was then used to fine-tune the registration of the functional data to the structural reference volume (freesurfer bbregister).

Then, the functional data was projected onto the extracted cortical surfaces by averaging between white and pial surfaces along the surface normal (mri\_vol2surf). To minimize partial volume effects, values were only averaged between 20% and 80% of the distance between the white and pial surfaces along the normal.

Finally, the functional data was smoothed along the surface with a Gaussian filter and then projected on to an icosahedron with uniform spacing of vertices in the spherical template space (freesurfer mri\_surf2surf). We experimented with several different smoothing values and icosahedron order numbers (which determines the density of vertices on the sphere, i.e., the resolution of the data) to determine parameters that sufficiently reduce the dimensionality of the data while still retaining as much information as possible.

## Machine Learning

ANOVA feature-selection (ANOVA ref and sk-learn ref)

Cross-validation (cv ref)

SVM (SVM ref)

SVR (SVR ref)

### Filtering

Block filtering (neurometrics1 ref)

#### Hidden Markov Model

To accurately model the state transition probabilities we found that we needed to use a discrete number of previous states. The set of brain states we are interested in decoding is , where for our stimulus. The domain of the hidden states is , where is the number of memory states we use. This results in an extremely large but sparse state transition matrix . The domain of the observable sequence is where is the number of voxels we are measuring. We use a trained feed-forward neural network to approximate from the observed fMRI data (Richard and Lippmann (1991)). We can’t generate an observation probability matrix because our observable domain is continuous, but by fitting a distribution to we can calculate using Bayes’ rule. Fortunately, prior fMRI studies have shown that the distribution of fMRI data can be reasonably approximated by a multivariate Gaussian after whitening (Worsley (2001)).

#### Modified Forward Algorithm

Based on this model, we propose the following modified forward algorithm for fMRI data. Given observations , a probability transition matrix , an observation probability distribution , a function that approximates , an initial state distribution , and , the filtered distribution for the state can be determined using Algorithm 4. The performance of this algorithm was estimated using a cross-validation procedure (Kohavi (1995)) because the distribution and the function need to be constructed on a training set. We compared its performance to simple block averaging on previously collected fMRI data.

**Algorithm 4** Forward Algorithm

### Within-Session

Compared volume, smoothed volume, surface, smoothed surface, and down-sampled icosahedron classification accuracy

### Within-Subject

Compared different smoothing and down-sampling icosahedron classification accuracy

### Across-Subject

Evaluated classification accuracy

## Sensitivity Analysis

# Results

## Psychophysics

Each subject with fits

## Volume/Smoothed Volume/Surface/Smoothed Surface/Sphere within-session comparison

Figure

## Within-Subject Classification/Regression

With and without filtering

## Spherical smoothing/downsampling average within-subject comparison

Figure with filtering

## Across-Subject Classification/Regression

With and without filtering

## Sensitivity Analysis

Average within-subject map and/or across-subject map

# Discussion

Discussion

# Bibliography

bib