Title

# Abstract

Abstract text

# Introduction

Decoding cognitive states (refs) from fMRI opens up exciting new applications in training and therapy. However, many hurdles remain to be cleared before real-world applications are possible: training or therapy must be performed within the scanner, decoding techniques must be able to handle less structured and controlled stimuli to accommodate training and therapy programs, we must be able to decode across different sessions to reduce total time in the scanner, and we must be able to decode variables that are inherent to the subject rather than the stimuli.

Subjects must remain still inside a cramped and noisy machine during their training or therapy session for the best decoding results. This severely limits the types of programs were fMRI decoding is realistically applicable. A potential solution to this limitation is to use virtual reality environments for the stimuli. While motion and other somatosensory inputs are still restricted, virtual environments provide a more immersive experience that should cause the subject's neural response to be closer to that of the real world. Encouragingly, some therapies already utilize virtual reality such as PTSD extinction therapy (refs).

Most cognitive neuroscience experiments are focused on explanatory power at the expense of classification accuracy and stimuli complexity. This makes sense when the objective is answering a basic research question. However, our goal is to apply fMRI cognitive state decoding during training and therapy where classification accuracy is extremely important and the stimuli will likely be highly complex. Therefore, we propose the use of more complex multivariate decoding techniques including feed-forward neural-networks and time-dependent filtering. These techniques are more difficult to interpret in terms of significant correlation between stimuli and response but provide significantly 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 attempt to solve this problem such as PCA (refs) and ICA (refs). 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. In this way, we leverage our knowledge of the structure of the brain to reduce the dimensionality of the time series data.

It’s important to be able to train the classifier in one session, and then use that trained classifier in subsequent sessions. This allows us to reduce the necessary scanning time per subject and also opens up the possibility of a decoding algorithm that gets better with each successive scan. Ideally, we would also like to be able to user decoders trained on a large number of different subjects. The main difficulty to solving both of these problems is accurately registering the volumes between different sessions and subjects. We have found that spherical registration (ref) with surface based smoothing and down-sampling yields the best results for both cross-session and cross-subject registration in terms of decoding accuracy. 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.

It is important that we decode a variable inherent to the subject rather than a reflection of the stimuli. Decoding the presentation of a particular visual or audio stimuli are uninteresting for our purposes; we already know what stimuli we are presenting to the subject. In the work presented here, we propose to decode the subject’s performance at a difficult task. This is something we don’t already know when we present the stimuli. 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.

# 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 environment very similar to those used in current PTSD extinction therapy programs (see figure #). Subjects perform a weapon detection task at 6 different locations in this virtual town. Between detection tasks, the screen smoothly transitions to a new location over 5 seconds, then a cue is presented to the subject that indicates the difficulty of the upcoming block for 5 seconds. Then the subject performs the task at that location for 45 seconds. During the task, 20 characters move through subject’s view, and a random character will pull out a weapon every 2—4 s. Subjects press a button to indicate they detected the weapon and the character puts the weapon away. If after a certain amount of time the subject fails to notice the weapon, the character puts the weapon away. After the task, we have a control period that begins with another 5-second cue followed by 15 seconds of the characters moving around and pulling out flashlights rather than weapons. The subject is instructed not to respond during this period and any responses are considered incorrect. The expected difficulty of the stimuli is adjusted each time the subject moves to a new location. However, the difficulty settings and locations were balanced so that there would be no correlation between location and task difficulty. Furthermore, low-level contrast was held constant in real-time using a post-processing shader.

|  |
| --- |
| Screenshot from stimulus of 1 location |
| Screenshot from stimulus of another location, in this image a character is holding a weapon that is circled. |

The duration that the character holds the weapon is dependent on the task difficulty and the actual visibility of the object. Since all of the characters are moving randomly and independently, the weapon frequently becomes partially or fully occluded during the search task. To account for this variability in difficulty, the duration that the weapon is visible is determined in real-time by its visibility. We calculate the total area of visible weapon in pixels and multiply this by the frame rate of the stimuli to arrive at the weapon’s total visibility in pixel-seconds. The character holds the weapon until the accumulated pixel-seconds exceed the threshold for the current difficulty setting. In addition to total visibility, varying the 3D depth of the character with the weapon also controlled task difficulty. The two variables are linked in that a weapon further away has a smaller projected area in pixels, so total visibility was calculated based on a constant factor from the distance.

## Psychophysics

Performance data was collected for each subject outside of the scanner. This data was used to individually tune the difficulty of the task to ensure that we would be able to collect an approximately uniform distribution of subject performance scores during the scanning session as well as to ensure that the subjects were not responding during the control period. Task performance was estimated as the average performance of the subject during a 30-second block while the distance of the weapon from the viewer and its total visibility were used to estimate difficulty. However, due to the semi-random movements of the characters, effective distance and visibility could vary considerably. Therefore, we collected extensive logging information during the stimuli that included the exact timing, positions, and visibility of all characters and weapons as well as the responses of the subject. Using this data, we fit sigmoid curves that allowed us to predict with reasonable accuracy the subject’s performance based on the difficulty setting of the stimuli (Figure #). During training periods outside of the scanner, the subject received audio feedback for successfully finding the weapon, missing a weapon, and indicating that they have seen a weapon while one is not present. However, during the scanning sessions there were no feedback cues.

## 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

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

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

Cross-validation (cv ref)

SVM (SVM ref)

### Filtering

Block filtering (neurometrics1 ref)

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

Figure

## Spherical smoothing/downsampling average within-subject comparison

Figure with filtering

## Within-subject Learning Curve Analysis and Accuracy

With and without filtering

## Across-Subject Learning Curve Analysis and Accuracy

With and without filtering

## Sensitivity Analysis

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

# Discussion

Discussion

# Bibliography

bib