ARO Progress Report

# Introduction

One of the long-term goals of this research is to create a tool that can assist in virtual reality training or therapy by providing measures of internal cognitive variables that are otherwise difficult to ascertain. In our previous study, we showed that it is possible to decode complex visual information from the pattern of BOLD activity in fMRI data while the subject views a virtual environment. This was an important first step to determine the feasibility of such a system, but visual information is clearly evident in the visual stimulus, and can be measured without fMRI.

Now, we are attempting to decode an internal cognitive variable: task difficulty. This is potentially a very useful measure as it would allow the difficulty of training or therapy tasks to be modulated to maximize subject attention even when the task lacks easily computed performance metrics. To accomplish this, we have designed a new stimulus, with substantially improved visual quality, upon which subjects perform a visual detection task of varying difficulty. Our previous study lacked an explicit task, and we believe this caused strong subject-to-subject and session-to-session variability. Therefore, we expect the addition of a task to help stabilize this variability in addition to giving us a far more relevant classification target.

Multi-variate pattern analysis (MVPA; Pereira, Mitchell, & Botvinick, 2009) is used to train classifiers that can estimate the difficulty of the task the subject is performing from the pattern of BOLD activation in the fMRI data, and the expected accuracy of the classifier on new data is estimated using statistical techniques. To gain some insight as to how the classifiers are accomplishing this task, we build sensitivity maps and project them onto the cortical surface for visualization. We have also begun experimenting with the application of intrinsic connectivity networks (ICNs; Laird et al., 2011) to help tease apart the contribution of various networks to the classifiers performance, and thereby provide a more semantic description of what’s going on.

# Methods

## Stimulus

|  |  |
| --- | --- |
| (a) | (b) |

Figure . Example frames from the stimulus during task (a) and rest (b) phases.

We built a new stimulus from the ground up using the Unity game engine (<http://unity3d.com>) in order to improve both the realism of the virtual environment and the degree of control over presentation. New high-resolution models and textures were used to greatly improve the visual quality of the environment over the previous iteration. Additionally, characters were given improved animations and programmed with basic artificial intelligence to increase the realism of the environment. Given our long-term interest in PTSD, the virtual environment is intended to reflect real-world environments encountered by our military personnel. It is similar to the virtual Iraq used in recent virtual reality exposure therapy experiments to treat PTSD (Gerardi, Rothbaum, Ressler, Heekin, & Rizzo, 2008).

The stimulus alternated between task periods and rest periods, which lasted for 30 seconds each. During the rest period, the subject’s perspective moved along fixed paths through the virtual town (Figure 1b). During the task period, the subject’s perspective was stationary, but a large group of characters randomly milled around the scene while avoiding collisions with each other (Figure 1a).

The subject was tasked with identifying when a character draws a firearm. Periodically (every 2–6 s), one of the characters milling around the scene was selected at random to draw a firearm and then holster it after a fixed period of time. The subject was instructed to press a button whenever he identified a target character with a drawn firearm. A correct response was recorded if the subject pressed the button during the presentation period and an incorrect response was recorded if the subject failed to press the button before the firearm was holstered or the subject pressed the button when no character had drawn a firearm. A correct button press caused the target character to immediately holster their weapon, providing feedback to the subject.

Difficulty of the task was be controlled by adjusting the type of firearm the characters drew (pistol, rifle, or rocket-propelled grenade launcher), the distance of the character from the subject’s perspective, and the duration that the firearm was presented for. Extensive subject response data and target visibility information was written out to an XML file during game play which allowed us to determine the relationship between these parameters and the subject’s probability of success. We found a considerable amount of variability in the subject’s performance was explained by what we call total visibility: the visible area of the object to be detected in pixels integrated in time across the duration of its presentation (with the resulting units being pixel-seconds). Therefore, we calculated total visibility in real time to adjust the presentation period of the firearm. This allowed us to compensate for the variance introduced by partial and total occlusion of the firearm caused by the randomly moving characters. To create a single difficulty control, we held the type of firearm constant and linked total visibility to presentation distance, i.e., for a difficult trial, the character would be presented further away and with a smaller total visibility.

We collected fMRI data using 3 different variants of the task. In the first variant, difficulty was held constant during each 30-second block and the subject received a cue indicating the difficulty of that block before it started. Difficulty was only controlled by distance and the firearm presentation period was fixed. In the second variant, difficulty varied continuously during the block and no cue was presented. Difficulty was controlled using both distance and total visibility to adjust the presentation period. In the third variant, difficulty was held constant during each block and the subject was cued. Difficulty was controlled using both distance and total visibility to adjust the presentation period.

The task and difficulty parameters were designed to limit the correlation between low-level visual cues and task difficulty. Although the distance of the character selected to draw a firearm varied with difficulty, the average distance, density, and speed of the characters remained the same in all difficulty settings. The character with the firearm moved and behaved exactly like the other characters. The only difference between the difficulty settings is the size and location of the firearm itself, which has a very small size and contrast compared to the entire scene.

## MRI

MRI data was collected on a Siemens Skyra 3T scanner. Compared to our previous work, this scanner allowed us to collect higher resolution data over a larger area of the brain without overheating; we could also collect more runs per session (see below). Four subjects participated in the experiments after giving informed consent under a human-subjects protocol approved by the UT Austin IRB.

### Anatomy

High-resolution anatomy scans were collected for each subject. Automatic cortical segmentation and surface extraction was performed on these volumes using the Freesurfer image analysis suite, which is documented and freely available online (<http://surfer.nmr.mgh.harvard.edu>).

### Functional

Functional data was collected using the Siemens product EPI sequence with grappa acceleration factor 3, 2 mm isometric voxels, 60 slices, and TR = 2.5 s. The prescription covered the whole brain and was oriented along the AC-PC axis. During each acquisition session, 8 runs were collected. Each run consisted of 8 alternations between 30-second task and rest periods for a total length of 6 minutes and 144 volumes. Motion correction and slice timing correction was performed using mrVista, which is documented and freely available online (<http://white.stanford.edu/software>).

### Resting State

Resting state data was collected using similar acquisition parameters. During each session, six 6-minute duration runs were collected. Motion correction and slice timing correction were again performed using mrVista.

ICA extraction was performed using the MELODIC tool from the FSL analysis suite (<http://fsl.fmrib.ox.ac.uk>). MELODIC was configured to extract 70 components based on the findings of Ray et al. (2013). The resulting independent components were analyzed by hand to determine whether they appeared to be physiological noise, motion artifacts, or ICNs. Those identified as ICNs were also given semantic labels as outlined in Laird et al. (2011).

## Machine Learning

Task difficulty was estimated for each frame using the target distance and total visibility information recorded during fMRI data acquisition along with the subject’s psychophysics curve. This continuous measure of task difficulty was then binned into 3 groups of equal size, i.e., each bin contained the same number of frames rather than being equal in width. We refer to these groups as easy, medium, and hard frames. These are the target classes which our machine learning algorithms were trained to classify.

We explored classifying these targets using whole-brain data as well as single ICNs. For single ICNs, masks were calculated by MELODIC and applied to the data before further processing. We selected 5 ICNs related to bottom-up attention and vision. We also selected 2 masks associated with physiological noise and motion artifacts as a sanity check.

Task relevant voxels were selected using an ANOVA approach (Scheffe, 1959). A single factor ANOVA analysis was performed independently on each voxel with the factor being whether the sample was recorded during a rest period or a task period. Voxels were then sorted based on their resulting F-test score and the first 3,000 voxels were selected for whole-brain data and the first 1,000 voxels were selected for ICNs. This served to reduce the dimensionality of the learning problem and thereby improve both training time and classifier performance.

The time series from these task relevant voxels and their associated difficulty targets were used to train a variety of machine learning algorithms. We found that the feed forward neural network (NN; Jain & Mohiuddin, 1996) had the best performance in this domain closely followed by the support vector machine (SVM; Cortes & Vapnik, 1995). Based on these findings, we will primarily focus on the results from the NN. Classifier accuracy was estimated using cross-validation (Kohavi, 1995). In our previous work we found that splitting the folds across the runs greatly reduced the bias of the performance estimate and therefore we used 8-fold cross-validation. P-values were estimated using a permutation test (Ojala & Garriga, 2010).

Sensitivity maps were constructed for each subject from neural networks trained on the whole-brain data and for each ICN (see our previous report; paper in preparation). These sensitivity maps were projected onto the Freesurfer generated cortical surfaces for visualization.

# Results

## Psychophysics

The aggregate psychophysics plots for each subject are presented in Figure 2. These measurements include total visibility correction. Confidence intervals were calculated using the Clopper-Pearson method for binomial random variables (Clopper & Pearson, 1934). Although some variability is present between subjects and within difficulty settings, all of the curves have a significant downward trend as the difficulty increases.

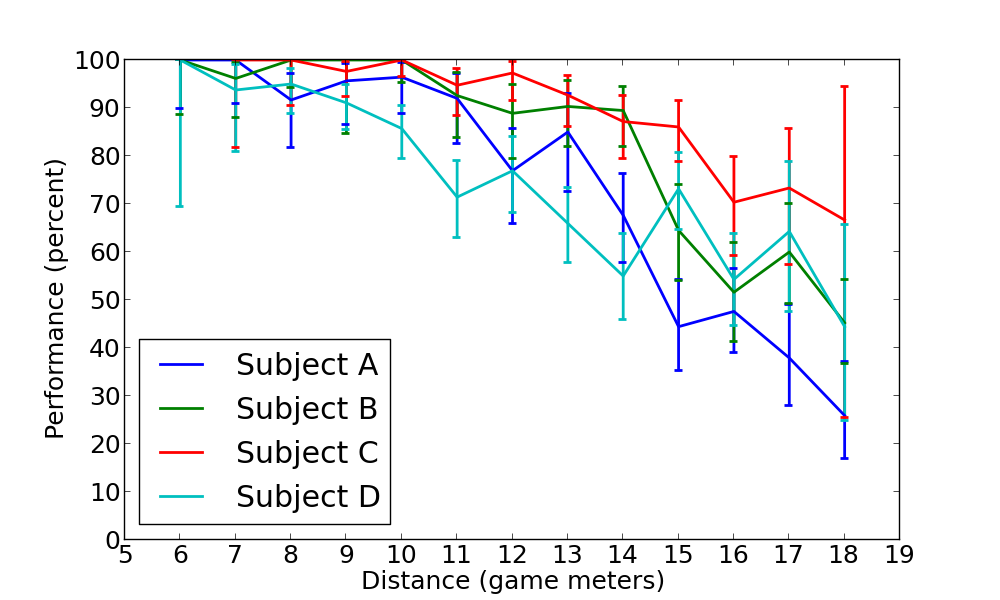


Figure . Aggregate psychophysics plots for each subject. These measurements include total visibility correction.

## Machine Learning

Individual cross-validated accuracy for the NN and SVM using task-relevant feature selection on whole-brain data is presented in Table 1. The average classification accuracy across all subjects and stimulus variants is 48%. The performance of all subjects is significantly above chance (p < 0.05).

Table . Individual cross-validated accuracy for the NN across all 3 stimulus variants.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Stimulus Variant | Cued, no visibility control | | Non-cued, visibility control | | Cued, visibility control | |
| Subject | A | D | A | B | C | D |
| NN | 51%\* | 51%\* | 47%\* | 43%\* | 50%\* | 47%\* |
| SVM | 47%\* | 52%\* | 44%\* | 40%\* | 38%\* | 41%\* |

Sensitivity maps from whole-brain data are presented in Figure 3. These maps show high levels of sensitivity in the temporoparietal junction (TPJ) and lower levels of sensitivity in visual areas and premotor cortex.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Left Lateral | Left Medial | Right Lateral | Right Medial |
| Subject A |  |  |  |  |
| Subject B |  |  |  |  |
| Subject C |  |  |  |  |
| Subject D |  |  |  |  |

Figure . Sensitivity maps from whole-brain data for each subject. The sensitivity maps have been projected onto the cortical surface.

So far, the ICN analysis has only been completed for subject C. The ICN masks used in this analysis are presented in Figure 4. ICN 4 corresponds to the bottom-up attention network; ICNs 5, 8, 10, and 21 correspond to different vision networks; ICN 7 corresponds to physiological noise; and ICN 11 corresponds to motion artifacts. The preliminary classification results from the analysis are presented in Table 2. Only ICNs 4 and 10—corresponding to the bottom-up attention network and one of the vision networks—had cross-validated accuracies significantly (p<0.05) above chance.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ICN | Left Lateral | Left Medial | Right Lateral | Right Medial |
| 4 |  |  |  |  |
| 5 |  |  |  |  |
| 7 |  |  |  |  |
| 8 |  |  |  |  |
| 10 |  |  |  |  |
| 11 |  |  |  |  |
| 21 |  |  |  |  |

Figure . Masks used in ICN analysis projected onto cortical surface.

Table . Cross-validated accuracy estimates for each ICN.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Subject C | | | | | | |
| ICN | 4 | 5 | 7 | 8 | 10 | 11 | 21 |
| NN | 43%\* | 35% | 36% | 36% | 40%\* | 35% | 37% |

The sensitivity maps for ICN 4 and 10 are presented in Figure 5. [Note areas highlighted, also need to double check figures since they look reversed.]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ICN | Left Lateral | Left Medial | Right Lateral | Right Medial |
| 4 |  |  |  |  |
| 10 |  |  |  |  |

Figure . Sensitivity maps for ICNs 4 and 10, corresponding to the bottom-up attention network and posterior visual network respectively. These were the only ICNs of those selected for analysis that had a cross-validated classification accuracy significantly above chance (p<0.05).

# Discussion and Ongoing Work

The performance results are promising and show that decoding an internal mental variable while a subject performs a task in a realistic virtual environment is feasible. Although the performance is lower than when we were decoding visual information, the results are far more interesting. Additionally, we expect to be able to improve the results by further polishing the stimulus and applying more advanced machine learning techniques.

In addition to controlling the visibility of the target, we are modifying the stimulus to better localize task-relevant activation for improved classification performance. Currently, we select voxels for further processing based on a single-factor ANOVA test between task and rest periods. However, there are significant visual differences in the stimulus between these two periods as well. To solve this problem, we are restructuring the stimulus to minimize visual differences between these periods. Additionally, we are experimenting with reducing the duration of the rest period to maximize the number of useful frames we collect each session.

We have preliminary evidence that suggests that combining classification outputs from each frame in a block to produce a single classification for the block significantly increases performance. The first combination scheme we tried was a simple voting mechanism, but we are exploring other more complex methods as well. We are also experimenting with training hierarchical SVMs and NNs on the ICN subsets as well as anatomic percellations. The hierarchy should help the machine learning algorithms exploit these inherent structures in the data. Due to their complexity, we are building a new computational framework to efficiently train and test these networks. We are also exploring the use of classification confidence measures. This is a machine learning topic that has received little attention, but may be very important in the training and therapy domain. Frequently when a machine learning algorithm misclassifies an input, it is because that input is either ambiguous or unlike any example the classifier was trained on. It would be helpful to have the classifier report some measure of confidence, particularly in the case of real-time feedback where the subject can be guided back to a state where the classifier is more reliable. Additionally, the trainer or therapist could ask the subject to self report only when the classifier reports low confidence. These data samples could then be used to retrain an even better classifier. We are performing a thorough investigation of the reliability of different confidence measures on a variety different machine learning algorithms to determine their effectiveness.

We are currently in the process of identifying additional internal cognitive variables to target. We are planning experiments where we encourage the subject to use different high-level strategies for completing tasks. In this way we would be measuring distinct neural states rather than single cognitive variables. We expect to see the different strategies reflected in the spatial distributions shown by the sensitivity maps and the ICNs contributing to classification. Another potential target is the subjective units of distress scale (SUDS). In the work of Gerardi, Rothbaum, Ressler, Heekin, & Rizzo (2008), the patient undergoing the virtual reality expose therapy is asked to report his current level of distress in SUDS every 10 seconds. Providing this same measure to the therapist based on patterns of neural activation without having to interrupt the immersion of the procedure would be extremely helpful. Although there are additional complications associated with measuring distress and anxiety, our results thus far indicate that such a system is within the realm of possibility.

# References

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 404–413.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. doi:10.1007/BF00994018

Gerardi, M., Rothbaum, B. O., Ressler, K., Heekin, M., & Rizzo, A. (2008). Virtual reality exposure therapy using a virtual Iraq: case report. *Journal of Traumatic Stress*, *21*(2), 209–13. doi:10.1002/jts.20331

Jain, A. K., & Mohiuddin, K. M. (1996). Artificial neural networks: a tutorial. *Computer*, *29*(3), 31–44. doi:10.1109/2.485891

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *International Joint Conference on Artificial Intelligence* (pp. 1137–1145).

Laird, A. R., Fox, P. M., Eickhoff, S. B., Turner, J. a, Ray, K. L., McKay, D. R., … Fox, P. T. (2011). Behavioral interpretations of intrinsic connectivity networks. *Journal of Cognitive Neuroscience*, *23*(12), 4022–37. doi:10.1162/jocn\_a\_00077

Ojala, M., & Garriga, G. (2010). Permutation tests for studying classifier performance. *The Journal of Machine Learning Research*, *11*, 1833–1863. Retrieved from http://dl.acm.org/citation.cfm?id=1859913

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage*, *45*(1 Suppl), S199–209. doi:10.1016/j.neuroimage.2008.11.007

Ray, K. L., McKay, D. R., Fox, P. M., Riedel, M. C., Uecker, A. M., Beckmann, C. F., … Laird, A. R. (2013). ICA model order selection of task co-activation networks. *Frontiers in Neuroscience*, *7*(December), 237. doi:10.3389/fnins.2013.00237

Scheffe, H. (1959). *The analysis of variance* (Vol. 72). John Wiley & Sons.