## Chapter 2: Background

Recent research has shown that fMRI is capable of decoding some cognitive states (Mitchell et al., 2004) such as the cognitive states associated with the perception of various types of objects (Cabral, Silveira, & Figueiredo, 2012; Shinkareva et al., 2008), what a person is saying and who is saying it (Formisano, De Martino, Bonte, & Goebel, 2008), and telling the truth or lying (Fan, Shen, & Davatzikos, 2006). The ability to decode cognitive states during training and therapy exercises could be invaluable for improving their efficacy. Virtual environments (VEs) are the most practical way to perform such exercises within the confines of an MRI scanner, and a number of virtual training and therapy environments already exist (Gerardi, Rothbaum, Ressler, Heekin, & Rizzo, 2008; Gonçalves, Pedrozo, Coutinho, Figueira, & Ventura, 2012). However, these exercises are far from the controlled stimuli used in most fMRI experiments. During such natural tasks, we expect a variety of complex interactions between many regions of the brain. A goal of the work reported in this paper is further development of computational analysis techniques that improve decoding accuracy of cognitive states in such an environment. Rather than focusing on a specific set of cognitive states, we look to develop a general approach to decoding task-relevant states with high accuracy. Additionally, the visual richness, the motion of the objects and viewer, and the real-time interaction with the virtual environment bring experiences to the subject much closer to those which shaped the evolution of our brains. It seems plausible then that using virtual environments can reveal how the brain functions under more realistic circumstances. Therefore, another goal of the work reported in this paper is to further development of analysis techniques that improve the interpretability of complex decoding algorithms for use in hypothesis driven experiments.

Using virtual environments in fMRI experiments have been explored by various researchers over the past decade. Early examples can be found in the work of (Spiers & Maguire, 2007b) and the resulting publications (Valente, De Martino, Esposito, Goebel, & Formisano, 2011) from the PBAIC 2007 competition (see <http://www.lrdc.pitt.edu/ebc/2007/competition.html>). In the case of Spiers & Maguire, a commercial taxi driving game was used as the stimulus, thereby leveraging many millions of dollars in development expense, but at the same time severely limiting control of the stimulus by the researchers. The game play was recorded during scanning, and afterwards the subject reviewed the video with a researcher and explained what they were thinking and doing at each point to assist in labeling the data. For the PBAIC 2007 competition, researchers constructed an interactive VE using the Source game engine. Subjects were given a relatively complex task to search for fruits, toy weapons, and characters with piercings, while avoiding contact with a dog. Subjects received compensation after the scan based on the score they received in the game. Similar to the work of Spiers & Maguire, the game play was recorded and participants rated their subjective mood along several axes, including arousal and valence, while reviewing the video. More recently, researchers have begun using VEs in more traditional controlled experimental protocols utilizing specially designed and far simpler VEs (Marsh et al., 2010; Mueller et al., 2012; Op de Beeck, Vermaercke, Woolley, & Wenderoth, 2013; Schindler & Bartels, 2013). These stimuli cost much less than a commercial game, and so are commensurately less realistic.

For our experiments, we used a virtual environment specially developed for us by a professional game and simulation designer using a state-of-the-art game engine. The visual quality of the environment and the motion of characters and camera were comparable to what is found in military training systems. The visual quality is similar to the stimuli used in the work of Spiers & Maguire as well as the PBAIC 2007 competition. However, in those stimuli the induced cognitive states are not well balanced. Due to their interactive nature, the subjects may spend significantly more time in one state than another. This complicates the training and, in particular, the evaluation of decoding algorithms. In our stimulus, we have balanced the induced states at the cost of interactivity to provide better accuracy estimates of different decoding methods for comparison. On the other hand, the stimulus is considerably more realistic – and the subject’s state less controlled – than what is found in the recent neuroscientific investigations involving VEs (Marsh et al., 2010; Mueller et al., 2012; Op de Beeck, Vermaercke, Woolley, & Wenderoth, 2013; Schindler & Bartels, 2013). It was important to measure the performance of different decoding methods in this environment to gauge their potential for use with training and therapy exercises.

Our goals were focused on exploring and improving methods of data analysis coupled with virtual environment stimulus design, rather than testing a specific neuroscience hypothesis. We aimed to extract the cognitive state of the subject associated with freely viewing a number of characters, rather than test the many possible perceptual mechanisms that encode this information in the human brain, such as object recognition, eye movements, or social group perception. Such decoding methods will be important for use of fMRI in clinical settings where it is useful to know the task-relevant cognitive state of the subject, but the neural mechanisms may not be well understood yet. We are, for example, interested in supporting work using virtual reality to treat PTSD due to combat, in which treatment exposes the subject to virtual stimuli that are highly suggestive of the physical situations that induced the trauma. Through carefully controlled use of VR, the patient is gradually desensitized over a period of weeks so that the likelihood of triggering of the trauma declines (Gerardi et al., 2008), as confirmed through fMRI measurements (Gonçalves et al., 2012). Our stimulus and experiments were developed with this in mind. In particular, we created a virtual town suggestive of the Middle East, and populated the town with a combination of U.S soldiers and foreign combatants (**Error! Reference source not found.**).

Most neuroscience experiments analyze their data using hypothesis-based statistical techniques, such as the general linear model (GLM), Such methods can be very effective only when a distinct and testable hypothesis is available. However, the closer the stimuli get to realistic experiences, as offered using VE, the more difficult it becomes to isolate a tractable hypothesis. Moreover, it is likely that the more complex VE stimuli will evoke a more broadly distributed cortical response that includes both low-level sensory and higher-level associative regions. The treatment of each voxel independently by GLM cannot capture the structure of multi-voxel responses reflecting the coordinated activity these widely distributed brain regions. For all these reasons, we employ multi-voxel pattern analysis (MVPA) based on machine learning, an approach introduced in (Haxby et al., 2001).

We offer a new combination of methods to decode and analyze VE stimulus information from fMRI data. Most fMRI applications of machine learning have shown discrimination between distinct object categories (Haxby et al., 2001; Pereira, Mitchell, & Botvinick, 2009). More recently the relationship between multiple objects has been explored (Baeck, Wagemans, & Op de Beeck, 2013). Here we demonstrate that the cognitive state associated with object number rather than object classification can be decoded from fMRI data. Specifically, the cognitive state associated with viewing a number of animated characters, varying from 1—6 can be decoded in a dynamically changing virtual environment with accuracy from 58—93% (chance is 16.7%). Such high classification accuracy has important potential for real-time fMRI based therapies that adjust the stimulus in response to brain activity.

To achieve this performance, we experimented with four machine learning algorithms. We were particularly interested in artificial neural networks (NN) and support vector machines (SVM). For completeness, we also tested a Gaussian naive Bayes classifier (GNB) (Duda & Hart, 1973), and k-nearest neighbor classifier (KNN). The SVM is the most commonly used machine-learning algorithm in MVPA analyses (Pereira et al., 2009). However, we found that NNs also produced very favorable results. Recent advances in NNs, such as deep learning (Hinton, Osindero, & Teh, 2006a) and convolutional networks, have been outperforming traditional SVMs in a variety of domains (Cireşan, Meier, Masci, & Schmidhuber, 2012). Before jumping to these advanced techniques, we wanted to explore the application of relatively simple feed-forward NNs on fMRI data, and we propose several methods for improving their classification performance.

MVPA classification performance can tell us to what degree the time-series data can be used to decode a target category, but we also want to know which voxels are encoding the desired stimulus information. The searchlight technique (Kriegeskorte, Goebel, & Bandettini, 2006) can be used in conjunction with any machine-learning algorithm to create a map, but it does not fully utilize the spatially distributed multivariate nature of the classifier. For SVMs, the absolute discriminative map (Formisano, De Martino, & Valente, 2008) has been used. However, the discriminative map is limited to SVM algorithms. We propose a new mapping method similar to the absolute discriminative map based on a technique called sensitivity analysis (Zurada, Malinowski, & Cloete, 1994). For an SVM, the method reduces to approximately the absolute discriminative map. However, the method is more general and has been adapted to the NN. The sensitivity analysis is used in several ways. First, we examine the use of the sensitivity results to train NNs on high-dimensional fMRI data with relatively few training examples. Second, we present a method for producing informative maps from trained networks based on sensitivity analysis. Finally, we develop a technique based on recursive feature elimination (Guyon, Weston, Barnhill, & Vapnik, 2002) to determine appropriate thresholds for these sensitivity maps. The recursive feature elimination technique also acts as a multivariate feature reduction technique that can improve decoding performance. A similar method was applied to SVMs in the work of (De Martino et al., 2008).

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.

Hidden Markov Models represent a system where there is a hidden or unobserv- able random variable that is linked to an observable variable. They allows us to calculate the probability distribution of the hidden state at some point in time given a sequence of observed variables. The model only requires two conditions on the hidden and observed variables. First, the dynamics of the hidden states obey the Markov property that the distribution of a hidden state at a particular point in time is conditionally independent given the value of the state at the previous point in time. Second, the observed variables must follow a similar property in that the distribution of an observed variable at a particular point in time must be conditionally independent given the value of the hidden state at that time. This model is particularly relevant for many engineering problems where we wish to estimate the value of some unknown variable given imperfect measurements but the dynamics of the variable are well known and can be mod- eled. There are a number of useful conditional probabilities to consider in this model and we will be examining three of them: the probability distribution of a state given all previous observations, the probability of a state given all previ- ous and future observations, and the probability of the entire sequence of states given the entire sequence of observed variables. We will also be introducing algorithms for calculating these probability distributions.

### Hidden Markov Model

A Hidden Markov Model can be formally defined as the following. Let be a Markov chain on the discrete domain with transition probability matrix where . Let be a series of random variables on the discrete domain with observation probability matrix where and each is conditionally independent given .

#### Filtering

Filtering is the problem of finding the probability distribution of a hidden state given all previous observations. Formally, we can define this distribution as

However, let us first consider how to calculate the full joint probability:

Note that and are defined directly by our model, and is the same as equation 2 at time . This implies some recursive algorithm exists for calculating the joint probability. To recover the conditional probability we calculate

##### Forward Algorithm

The forward algorithm is an efficient solution to the filtering problem making use of the recursion identified in the previous equations. Given observations , a probability transition matrix , an observation probability matrix , an initial state distribution , and , the filtered distribution of state can be determined using Algorithm 1.

**Algorithm 1** Forward Algorithm

#### Smoothing

Smoothing is the problem of finding the probability distribution of a hidden state given all observations including those in the future. Unlike filtering, smoothing can not be performed in real-time as it requires future observations. Formally we can define this distribution as