

Learning and Predicting Brain Dynamics from fMRI: a Spectral Approach *

François Meyer^a

^aCenter for the Study of Brain, Mind and Behavior ;
Program in Applied and Computational Mathematics,
Princeton University

ABSTRACT

Traditional neuroimaging experiments, dictated by the dogma of functional specialization, aim at identifying regions of the brain that are maximally correlated with a simple cognitive or sensory stimulus. Very recently, functional MRI (fMRI) has been used to infer subjective experience and brain states of subjects immersed in natural environments. These environments are rich with uncontrolled stimuli and resemble real life experiences. Conventional methods of analysis of neuroimaging data fail to unravel the complex activity that natural environments elicit. The contribution of this work is a novel method to predict action and sensory experiences of a subject from fMRI. This method relies on an embedding that provides an optimal coordinate system to reduce the dimensionality of the fMRI dataset while preserving its intrinsic dynamics. We learn a set of time series that are implicit functions of the fMRI data, and predict the values of these time series in the future from the knowledge of the fMRI data only. We conducted several experiments with the datasets of the 2007 Pittsburgh Experience Based Cognition competition.

Keywords: fMRI, Laplacian eigenmaps, embedding, brain states.

1. INTRODUCTION

1.1 From localization to decoding

Functional MRI (fMRI) provides very coarse scale (as compared to the scale of neurons) measurements of brain activity. Traditionally the goal of neuroimaging has been to identify regions of the brain that are maximally correlated with a simple cognitive or sensory stimulus. This type of experimental paradigm is dictated by the dogma of functional specialization: cognitive and sensory processing is spatially localized. Very recently, fMRI has been used to infer subjective experience and brain states of subjects immersed in natural environments.¹⁻³ Such environments resemble real life situations with a wide variety of stimuli that are not controlled. Conventional univariate methods of analysis cannot be directly applied since there is no clearly identified stimulus. In⁴ the authors have asked subjects to watch a movie while in the scanner. Time series of the stimuli of interest (faces, voices, colors, etc.) were estimated independently by labeling the movie. More challenging experiments involve virtual reality environments.⁵ Content analysis and verbal reports were used to produce time series of potential stimuli to detect activation. Exploratory techniques, such as ICA and MVPA⁶ have also been used to disentangle the responses to external stimuli from the internal intrinsic variations.³

Despite the limited temporal and spatial resolutions of fMRI, there is considerable interest in trying to “decode” cognitive states from fMRI data.⁷ Practical experiments include decoding the contents of visual imagery.⁸ Decoding methods include linear discriminant analysis and support vector machines.⁹ A well defined experiment is provided by the Pittsburgh Brain Activity Interpretation Competition, where the goal is to learn and predict subject-driven actions and sensory experience from fMRI data associated with dynamic experiences in a virtual reality environment.¹⁰

This paper appeared in the proceedings of Wavelet XII conference, Proceedings of SPIE Volume 6701, 2007.
Corresponding author: fmeyer@Princeton.Edu, www.princeton.edu/~fmeyer

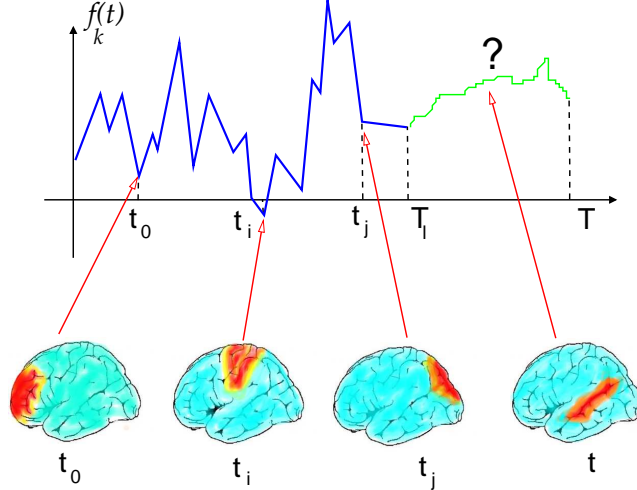


Figure 1. We study the variation of the set of features $f_k(t), k = 1, \dots, K$ as a function of the dynamical changes in the fMRI signal $\mathbf{X}(t) = [x_1(t), \dots, x_N(t)]$. We predict the feature functions f_k for $t = T_{l+1}, \dots, T$, from the knowledge of the entire fMRI dataset \mathcal{X} , and the partial knowledge of $f_k(t)$ for $t = 1, \dots, T_l$. The “toy” activation patterns illustrate the changes in “brain states” occurring as a function of time.

1.2 The EBC dataset

Several subjects were scanned while being immersed in a virtual reality environment. The goal of the competition is to learn the association between fMRI data and the subject’s mental states on a training set, and predict the mental states from fMRI only. Three subjects were scanned during three 20-minute segments (704 scanned samples in each segment). Time series of natural stimuli, or features, (25 in total) were recorded. Objective features included the presence of human faces, the presence of a dog, etc. Self-reported, and more subjective, features included the experience of happiness and fear. Subjects were instructed to complete various search tasks in the environment. Data were collected with a 3T EPI scanner. The feature time series were provided for the first two segments (1408 time samples) and competitive entries were judged on their ability to predict the feature on the third segment (704 time samples, see Fig. 1).

The contribution of this work is a novel method to predict action and sensory experiences of a subject from fMRI. We construct an embedding that maps each fMRI dataset to a lower dimensional space, replacing the voxels with an optimal coordinate system that preserves the dynamics of the fMRI signal. We learn the set of 25 features, which become implicit functions of the fMRI data, and we predict the values of these times series in the future from the knowledge of the fMRI data only.

2. THE MANIFOLD OF GLOBAL BRAIN STATES

2.1 A voxel-free parametrization of brain states

At a microscopic level, a large number of internal variables associated with various physical and physiological phenomena contribute to the dynamic changes in the fMRI signal. Because the fMRI signal is a large scale measurement of neuronal activity, we expect that many of these variables will be coupled resulting in a low dimensional set for all possible configurations of the activated fMRI signal. In this work we seek a low dimensional representation of the entire fMRI dataset that provides a new set of “voxel-free” coordinates to study cognitive and sensory features.

We denote a three-dimensional volumes of fMRI composed of a total of N voxels by $\mathbf{X}(t) = [x_1(t), \dots, x_N(t)]$. We have access to T such volumes. We can stack the spatio-temporal fMRI dataset into a $N \times T$ matrix,

$$\mathcal{X} = \begin{bmatrix} x_1(1) & \cdots & x_1(T) \\ \vdots & \vdots & \vdots \\ x_N(1) & \cdots & x_N(T) \end{bmatrix}, \quad (1)$$

where each row n represents a time series \mathbf{x}_n generated from voxel n and each column j represents a scan acquired at

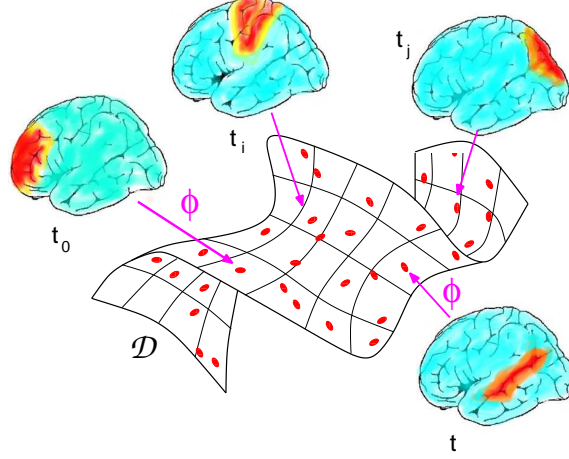


Figure 2. Low-dimensional parametrization of the set of “brain states”. The parametrization is constructed from the samples provided by the fMRI data at different times, and in different states.

time t_j . We call the set of features to be predicted $f_k, k = 1, \dots, K$. We are interested in studying the variation of the set of features $f_k(t), k = 1, \dots, K$ describing the subject experience as a function of the dynamical changes of the brain, as measured by $\mathbf{X}(t)$. Formally, we need to build predictions of $f_k(t)$ for $t = T_{l+1}, \dots, T$, from the knowledge of the entire fMRI dataset \mathcal{X} , and the partial knowledge of $f_k(t)$ for the training time samples $t = 1, \dots, T_l$ (see Fig. 1). We use here the global information provided by the dynamical evolution of $\mathbf{X}(t)$ over time, both during the training times and the test times. We would like to effectively replace each fMRI dataset $\mathbf{X}(t)$ by a small set of features that facilitates the identification of the brain states, and make the prediction of the features easier. Formally, our goal is to construct a map ϕ from the voxel space to a low dimensional space,

$$\phi : \mathbb{R}^N \mapsto \mathcal{D} \subset \mathbb{R}^L \quad (2)$$

$$\mathbf{X}(t) = [x_1(t), \dots, x_N(t)]^T \mapsto (y_1(t), \dots, y_L(t)), \quad (3)$$

where $L \ll N$. As t varies over the training and the test sets, we hope that we explore most of the possible brain configurations that are useful for predicting the features. The map ϕ provides a parametrization of the brain states. Figure 2 provides a pictorial rendition of the map ϕ . The range \mathcal{D} , represented in Fig. 2 as a smooth surface, is the set of parameters y_1, \dots, y_L that characterize the brain dynamics. Different values of the parameters produce different “brain states”, associated with different patterns of activation. Note that time does not play any role on \mathcal{D} , and neighboring points on \mathcal{D} correspond to similar brain states. Equipped with this re-parametrization of the dataset \mathcal{X} , the goal is to learn the evolution of the feature time series as a function of the new coordinates $[y_1(t), \dots, y_L(t)]^T$. Each feature function is an implicit function of the brain state measured by $[y_1(t), \dots, y_L(t)]$. For a given feature f_k , the training data provide us with samples of f_k at certain locations in \mathcal{D} .

The map ϕ is build by globally computing a new parametrization of the set $\{\mathbf{X}(1), \dots, \mathbf{X}(T)\}$. This parametrization is built into two stages. First, we construct a graph that is a proxy for the entire set of fMRI data $\{\mathbf{X}(1), \dots, \mathbf{X}(T)\}$. Second, we compute some eigenfunctions ϕ_k defined on the graph. Each eigenfunctions provides one specific coordinate for each node of the graph.

2.2 The graph of brain states

We represent the fMRI dataset for the training times and test times by a graph G that is constructed as follows. The 3-D fMRI volume at time t_i , $\mathbf{X}(t_i)$, becomes the node (or vertex) i of the graph. The weight $w_{i,j}$ on the edge $\{i, j\}$ quantifies dynamical changes between t_i and t_j . We found that measuring the normalized correlation between $\mathbf{X}(t_i)$ and $\mathbf{X}(t_j)$,

$$w_{i,j} = 1 - \frac{\langle \mathbf{X}(t_i), \mathbf{X}(t_j) \rangle}{\|\mathbf{X}(t_i)\| \|\mathbf{X}(t_j)\|},$$

is a good metric to build the graph. This metric is not sensitive to global changes of the signal (created by global blood flow changes, or residual motion, etc). It is much less sensitive to global changes in the intensity (e.g. residual motion) than

the Euclidean distance. Note that the distance currently compares all the voxels (white and gray matter, as well as CSF) inside the brain. The Pearson correlation coefficient was also used in^{4,11} to quantify changes in the fMRI signal triggered by natural stimuli. Finally, each node i is connected to its n_n nearest neighbors.

2.3 A new distance on the graph

Once the network of connected brain states is created, we need to define a distance between any two time samples t_i and t_j . This distance should reflect the topology of the graph, but should also be able to distinguish between strongly connected states (the two fMRI data are in the same cognitive state) and weakly connected states (the fMRI data are similar, but do not correspond to the same brain states). As shown,¹² the geodesic (shortest) distance between two nodes measured along the edges of the graph is very sensitive to short-circuits created by the noise in the fMRI data. A standard alternative to the geodesic distance is the commute time, $\kappa(i, j)$, that quantifies the expected path length between i and j for a random walk started at i .¹³ Let us consider a random walk Z_n on the connectivity graph. The walk starts at i , and evolves on the graph with the transition probability \mathbf{P} , defined by

$$P_{i,j} = w_{i,j}/d_i, \quad (4)$$

where $d_i = \sum_j w_{i,j}$ is the degree of the vertex i . We define \mathbf{D} to be the diagonal degree matrix with entries

$$D_{i,i} = d_i = \sum_j W_{i,j}. \quad (5)$$

If the walk is at i , it jumps to j (one of the nearest neighbors of i) with probability $P_{i,j}$. If t_i and t_j correspond to the same mental state, then the walk will jump from i to j with a high probability. Indeed, in this case $\mathbf{X}(t_i)$ and $\mathbf{X}(t_j)$ will be similar, and we will have $P_{j,k} \approx 1$. The average hitting time $H(i, j)$ measures the expected number of steps that it takes for the random walk Z_n started at i to hit j for the first time,¹³

$$H(i, j) = E_i[T_j] \quad \text{with} \quad T_j = \min\{n \geq 0; Z_n = j\}.$$

The commute time κ is simply a symmetric version of H ,¹³

$$\kappa(i, j) = H(i, j) + H(j, i) = E_i[T_j] + E_j[T_i]. \quad (6)$$

As one would expect, $\kappa(i, j)$ increases with the geodesic distance $\delta(i, j)$. Unlike the geodesic distance, $\kappa(i, j)$ decreases when the number of paths between the nodes increases. The reciprocal is not true: $\kappa(i, j)$ can be large even when $\delta(i, j)$ is small. The following double inequality compares the commute time and the geodesic distance.

THEOREM 2.1. *If i and j are at a distance $\delta(i, j)$ on the graph, then*

$$2\delta(i, j) \leq \kappa(i, j) \leq C\delta(i, j),$$

with $C = \max_{i,j} 1/(\pi_i P_{i,j}) = \sum_{i,j} w_{i,j} / \min_{i,j} w_{i,j}$. $\mathbf{\Pi}^T = [\pi_1 | \dots | \pi_T]$ is the stationary distribution associated with \mathbf{P} , $\mathbf{\Pi}^T \mathbf{P} = \mathbf{\Pi}^T$. The lower bound of (2.1.) is obvious. The upper bound is a direct consequence of the following result.

LEMMA 2.2. *If i and j are two adjacent vertices, then*

$$\kappa(i, j) \leq \frac{1}{\pi_i P_{i,j}}$$

Proof. From the strong Markov property,¹³ we have

$$\kappa(i, j) = \frac{1}{\pi_i P_i(T_i < T_j)}.$$

But if i and j are adjacent, the probability that, starting at i , j is visited before returning to i , $P(T_j < T_i)$, is greater than the probability that, starting from i , the walk visits j at the next instant. So,

$$\kappa(i, j) = \frac{1}{\pi_i P_i(T_i < T_j)} \leq \frac{1}{\pi_i P_{i,j}}.$$

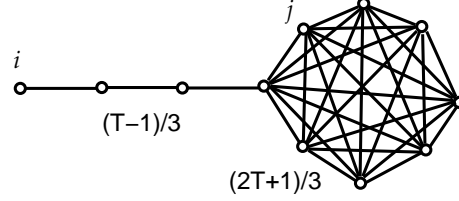


Figure 3. For i and j we have $\delta(i, j) = \frac{2}{3}T$, and $C = 2T(2T + 1)/18$.

□

The constant C can be quite large. Interestingly, among all graphs with T vertices, the lollipop graph (a path with $(T - 1)/3$ vertices attached to a complete subgraph with $(2T + 1)/3$ vertices) is the graph with the largest $\kappa(i, j)$,

THEOREM 2.3.¹⁴ *The commute time $\kappa(i, j)$ of the lollipop graph (see Fig. 3) is given by*

$$\kappa(i, j) = \frac{4}{27}T^3 + O(T).$$

For our analysis, the situation of Fig. 3 corresponds to a very interesting case: the fully connected sub-graph (clique) can be interpreted as a strongly connected set of brain states. The path, or handle of the lollipop, is interpreted as weakly connected brain states: the fMRI datasets are similar (small geodesic distance), but the corresponding brain states are very different. As opposed to the geodesic distance, the commute time between the brain states i and j remains very large, indicating that the states i and j are significantly different. We advocate that commute time provides a more robust measure of distance on the dataset.

The hitting time satisfies the following one-step equation,

$$E_i[T_j] = 1 + \sum_{k:k \neq j} P_{i,k} E_i[T_k]. \quad (7)$$

As a result, the hitting time can be expressed as a function of the fundamental matrix,¹³

$$\mathbf{Z} = (\mathbf{I} - (\mathbf{P} - \mathbf{\Pi}))^{-1} = \mathbf{I} + \sum_{k \geq 1} \mathbf{P}^k - \mathbf{\Pi}.$$

We have,

THEOREM 2.4.¹³ $E_i[T_j] = (Z_{j,j} - Z_{i,j})/\pi_j$.

Note that \mathbf{Z} is the Green function of the Laplacian. We can consider the eigenvectors ϕ_1, \dots, ϕ_T of

$$\mathbf{D}^{\frac{1}{2}} \mathbf{P} \mathbf{D}^{-\frac{1}{2}}, \quad (8)$$

which is a symmetric version of the transition matrix \mathbf{P} . The corresponding eigenvalues are between -1 and 1 , and can be labeled such that $-1 \leq \lambda_T \dots \leq \lambda_2 < \lambda_1 = 1$. Each eigenvector ϕ_k is a vector with T coordinates, one for each node of the graph G . We can therefore think of ϕ_k as being a function defined on the graph (see Fig 4). We will therefore write

$$\phi_k = [\phi_k(1), \phi_k(2), \dots, \phi_k(T)]^T \quad (9)$$

to emphasize the fact that ϕ_k is as a function defined on the nodes of the graph (see Fig 4).

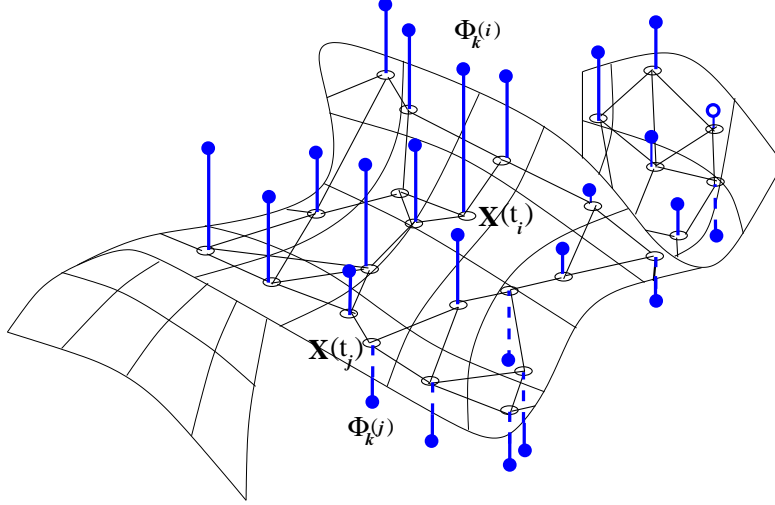


Figure 4. The eigenvector ϕ_k is a vector of T components. We plot the magnitude of each component, $\phi_k(i)$, as a blue stick: the stick points up if $\phi_k(i)$ is positive, down otherwise. We display the stick at the node \mathbf{x}_i corresponding to the i^{th} component $\phi_k(i)$. We can now think of ϕ_k as a function defined on the nodes of the graph.

Finally, from (2.4) the commute time can be expressed as

$$\kappa(i, j) = \sum_{k=2}^T \frac{1}{1 - \lambda_k} \left(\frac{\phi_k(i)}{\sqrt{\pi_i}} - \frac{\phi_k(j)}{\sqrt{\pi_j}} \right)^2, \quad (10)$$

The right hand side of (10) is the sum of $T - 1$ squared contributions of the form

$$\frac{1}{1 - \lambda_k} \left(\frac{\phi_k(i)}{\sqrt{\pi_i}} - \frac{\phi_k(j)}{\sqrt{\pi_j}} \right)^2.$$

In the expression above, we can associate the term $\phi_k(i)/\sqrt{\pi_i}$ with the node i , while $\phi_k(j)/\sqrt{\pi_j}$ should be associated with the node j .

2.4 Embedding of the dataset

We can then define a mapping from each fMRI scan $\mathbf{X}(t_i)$ to a vector of $T - 1$ coordinates,

$$\mathbf{X}(t_i) \mapsto \frac{1}{\sqrt{\pi_i}} \left[\frac{\phi_2(i)}{\sqrt{1 - \lambda_2}}, \dots, \frac{\phi_T(i)}{\sqrt{1 - \lambda_T}} \right]^T. \quad (11)$$

The idea was first proposed in¹⁵ to embed manifolds. Recently the same idea has been revisited in the machine learning literature.^{16,17} The approach has also been used to construct maps of spectral coherence of fMRI data in.¹⁸ According to (10), the commute time is simply the Euclidean distance measured between the new coordinates. In practice, we need not use all the eigenvectors in (11). Indeed, because $-1 \leq \lambda_T \leq \dots \leq \lambda_2 < \lambda_1 = 1$ we have

$$\frac{1}{1 - \lambda_2} > \frac{1}{1 - \lambda_3} > \dots > \frac{1}{1 - \lambda_T}.$$

We can therefore neglect $\phi_k(j)/\sqrt{1 - \lambda_k}$ for large k , and reduce the dimensionality of the embedding by using only the first L coordinates. We define the map ϕ from \mathbb{R}^T to \mathbb{R}^L ,

$$\mathbf{x}_i \mapsto \phi(\mathbf{x}_i) = \frac{1}{\sqrt{\pi_i}} \left[\frac{\phi_2(i)}{\sqrt{1 - \lambda_2}}, \dots, \frac{\phi_{L+1}(i)}{\sqrt{1 - \lambda_{L+1}}} \right]^T. \quad (12)$$

The spectral gap measures the difference between the first two eigenvalues, $\lambda_1 - \lambda_2 = 1 - \lambda_2$. A large spectral gap indicates that the low dimensional parametrization (12) will provide a good approximation. The construction of the embedding is summarized in Fig. 5.

Algorithm 1: Construction of the embedding

Input:

- $\mathbf{X}(t_i), i = 0, \dots, T - 1$,
- n_n number of nearest neighbors; L : number of eigenvectors.

Algorithm:

1. construct the graph defined by the n_n nearest neighbors of each $\mathbf{X}(t_i)$
2. compute \mathbf{P} and \mathbf{D} . Find the first L eigenvectors, ϕ_k , of $\mathbf{D}^{\frac{1}{2}} \mathbf{P} \mathbf{D}^{-\frac{1}{2}}$

Output: For all i

- new coordinates of $\mathbf{X}(t_i)$ given by (12):

$$[y_1(t_i), \dots, y_L(t_i)]^T = \frac{1}{\sqrt{\pi_i}} \left[\frac{\phi_2(i)}{\sqrt{1 - \lambda_2}}, \dots, \frac{\phi_{L+1}(i)}{\sqrt{1 - \lambda_{L+1}}} \right]^T.$$

Figure 5. Construction of the embedding

Unlike PCA which gives a set of basis vectors on which to project the dataset, the embedding (12) directly provides the new coordinates for each fMRI scan $\mathbf{X}(t_i)$. The new coordinates of $\mathbf{X}(t_i)$ are given by concatenating the i^{th} coordinates of the first L eigenvectors ϕ_k . Note that the first eigenvector ϕ_1 is not included as a coordinate because it is constant over all nodes of the graph.

2.5 How many new coordinates do we need ?

A parameter of the embedding (12) is L , the number of coordinates. L can be optimized by minimizing the prediction error. We expect that for small values of L the embedding will not describe the data with enough precision, and the prediction will be inaccurate. If L is too large, some of the new coordinates will be describing the noise in the dataset, and the algorithm will overfit the training data.

Fig. 6 illustrates the effect of L on the performance of the nonlinear dimension reduction. The quality of the prediction for the features: faces, instruction and velocity is plotted against L . Instructions elicits a strong response in the auditory cortex. Faces requires many more eigenfunctions to become optimal. The performance eventually becomes constant, and drops slightly for instructions. This confirms our hypothesis that we can replace 15,000 voxels with 50 appropriately chosen coordinates.

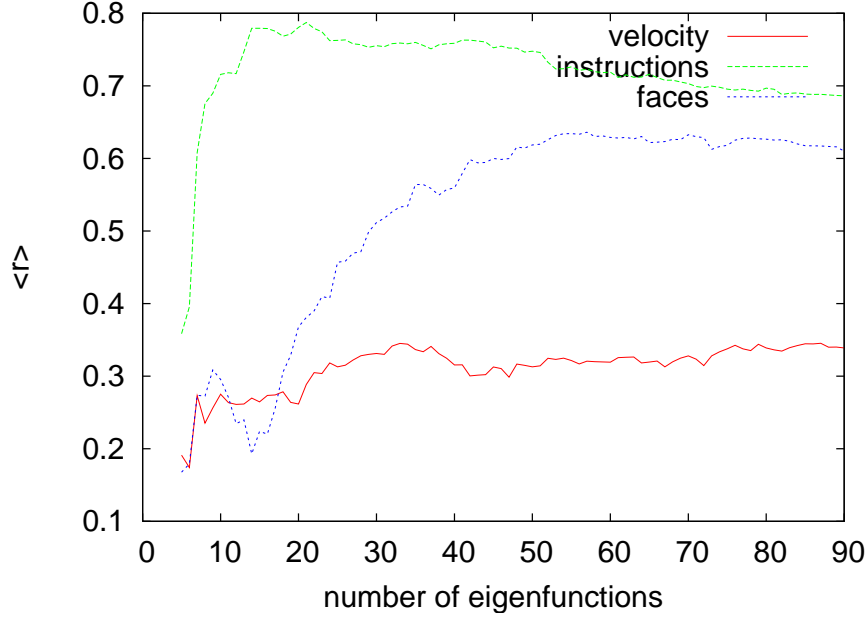


Figure 6. Quality of the prediction for three features: Faces, Instructions and Velocity as a function of the number L of coordinates.

3. SEMI-SUPERVISED LEARNING OF THE FEATURES

3.1 Kernel ridge regression

The problem of predicting a feature f_k at an unknown time t_u is formulated as kernel ridge regression problem.¹⁹ The training set $\{f_k(t)$ for $t = 1, \dots, T_l\}$ is used to estimate the optimal choice of weights in the following model,

$$\hat{f}(t_u) = \sum_{t=1}^{T_l} \hat{\alpha}(t) \mathcal{K}(y(t_u), y(t)),$$

where \mathcal{K} is a kernel (we use a Gaussian kernel in this work), and t_u is a time point where we need to predict.

3.2 Using the training data to construct the embedding

We have noticed that the construction of the graph could be improved by including the knowledge about the features $\{f_k(t)$ for $t = 1, \dots, T_l\}$ during the training time. In principle, the graph should only describe the manifold of brain states as measured by fMRI. However, the features measured during the training times can be used to penalize association of mental states that appear unlikely.

4. RESULTS

4.1 The eigen modes of the brain dynamics

We compared the nonlinear approach to dimension reduction to a linear analysis obtained with a principal components analysis (PCA) of \mathcal{X} . Here the principal components are principal volumes, and for each time t we can expand $\mathbf{X}(t)$ onto the principal components.

The table below summarizes the results. The 1408 training data were divided into two subsets of 704 time samples. We use $f_k(t)$ on the first part of the data ($1 \leq t \leq 704$) to predict $f_k(t)$ on the second part ($705 \leq t \leq 1408$). We then reverse the role of the training and testing sets. The performance is quantified with the normalized correlation between the model prediction and the real value of f_k ,

$$r = \langle \delta f_k^{est}(t), \delta f_k(t) \rangle / \sqrt{\langle \delta f_k^{est} \rangle^2 \langle \delta f_k^2 \rangle}, \quad (13)$$

where $\delta f_k = f_k(t) - \langle f_k \rangle$. Finally, r is averaged over the two training sets. The column κ -spec summarizes the performance of the approach based on the spectral decomposition of the commute-time, and summarized in Fig. 5. The column PCA + κ -spec displays the prediction error when a small number of PCA components were removed from the dataset before computing the optimal embedding. On average this pre-processing improved the performance by removing confounding effects. For a small number of features (e.g. fearful/anxious, instructions) this procedure decreased the performance of the prediction. For almost all the features the nonlinear approach performs better (r closer to 1) than PCA. These results were obtained with $K = 35$ for both approaches.

subject 14				subject 14			
feature	PCA	κ -spec	PCA + κ -spec	feature	PCA	κ -spec	PCA + κ -spec
AnnoyedAngry	-0.01	0.021	0.05	HitsWeapons	-0.03	0.17	0.18
Arousal	0.04	0.09	0.12	Hits	0.15	0.35	0.41
Body	0.07	0.31	0.45	Instructions	0.68	0.70	0.67
Dog	0.02	0.27	0.37	InteriorExterior	0.16	0.21	0.12
Faces	0.50	0.61	0.65	SearchPeople	0.16	-0.03	-0.02
FearfulAnxious	0.30	0.27	0.24	Velocity	0.19	0.34	0.47
HitsPeople	0.23	0.19	0.17				

4.2 Acknowledgments

The author is very grateful to the entire Princeton EBC team, Michael Bannert, Denis Chigirev, Jon Cohen, Greg Detre, Chris Moore, Ken Norman, Per Sederberg, Greg Stephens, Damon Tomlin, Matt Weber, and David Weiss, for a productive scientific interaction. In particular, the author thanks Per Sederberg and Greg Stephens for insightful discussions.

REFERENCES

1. Y. Golland, S. Bentin, H. Gelbard, Y. Benjamini, R. Heller, and Y. Nir *et al.*, “Extrinsic and intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation,” *Cerebral Cortex* **17**, pp. 766–777, 2007.
2. S. Malinen, Y. Hlushchuk, and R. Hari, “Towards natural stimulation in fMRI—issues of data analysis,” *NeuroImage* **35**, pp. 131–139, 2007.
3. H. Spiers and E. Maguire, “Decoding human brain activity during real-world experiences,” *Trends in cognitive sciences* **11(8)**, pp. 356–365, 2007.
4. A. Bartels and S. Zeki, “Functional brain mapping during free viewing of natural scenes,” *Human Brain Mapping* **21**, pp. 75–85, 2004.
5. H. Spiers and E. Maguire, “Thoughts, behaviour, and brain dynamics during navigation in the real world,” *Neuroimage* **31**, pp. 1826–1840, 2006.
6. K. Norman, S. Polyn, G. Detre, and J. Haxby, “Beyond mind-reading: multi-voxel pattern analysis of fmri data,” *Trends in cognitive sciences* **10(9)**, pp. 424–430, 2006.
7. J.-D. Haynes and G. Rees, “Decoding mental states from brain activity in humans,” *Nature Neuroscience* **7**, Apr. 2006.
8. J. Haxby, M. Gobbini, M. Furey, A. Ishai, J. Schouten, and P. Pietrini, “Distributed and overlapping representations of faces and objects in ventral temporal cortex,” *Science* **9**, pp. 2425–2430, 2001.
9. D. Cox and R. Savoy, “Functional magnetic resonance imaging (fmri) ”brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex,” *NeuroImage* **19**, pp. 261–270, 2003.
10. “<http://www.ebc.pitt.edu>.”

11. U. Hason et al., “Intersubject synchronization of cortical activity during natural vision,” *Science* **303**, pp. 1634–1640, 2004.
12. F. Meyer and X. Shen, “Exploration of high dimensional biomedical datasets with low-distortion embedding,” in *Proc. Data Mining for Biomedical Informatics Workshop, 7th SIAM International Conference on Data Mining*, 2007. *To appear*.
13. P. Bremaud, *Markov Chains*, Springer Verlag, 1999.
14. J. Jonasson, “Lollipop graphs are extremal for commute times,” *Random Structures and Algorithms* **16**(2), pp. 131–142, 2000.
15. P. Bérard, G. Besson, and S. Gallot, “Embeddings Riemannian manifolds by their heat kernel,” *Geometric and Functional Analysis* **4**(4), pp. 373–398, 1994.
16. M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neural Computations* **15**, pp. 1373–1396, 2003.
17. R. Coifman and S. Lafon, “Diffusion maps,” *Applied and Computational Harmonic Analysis* **21**, pp. 5–30, 2006.
18. B. Thirion, S. Dodel, and J.-B. Poline, “Detection of signal synchronizations in resting-state fMRI datasets,” *Neuroimage* **2**, pp. 321–27, 2006.
19. T. Hastie, R. Tibshinari, and J. Freedman, *The elements of statistical learning*, Springer Verlag, 2002.