COLLABORATIVE DENOISING OF MULTI-SUBJECT FMRI DATA

Alexander Lorbert⋆ *J. Swaroop Guntupalli*† *David J. Eis*⋆ *James V. Haxby*† *Peter J. Ramadge*⋆

⋆ Dept. of Electrical Engineering, Princeton University, Princeton NJ
† Center for Cognitive Neuroscience, Dartmouth College, Hanover NH

ABSTRACT

We propose a novel collaborative denoising scheme for multi-subject fMRI data. The scheme assumes that subjects experience a common, synchronous stimulus and uses the across-subject shared response structure to jointly denoise each subject's fMRI response along the spatial or voxel domain. Denoising is accomplished by learning subject-specfic orthonormal bases that yield sparse representations in a common transform domain. We provide empirical results using a real-world, multi-subject fMRI dataset.

***Index Terms*—** fMRI, Procrustes problems, signal denoising, principal axes

1. INTRODUCTION

A crucial step in the processing chain of fMRI data analysis is signal denoising. Denoising fMRI data is typically done (i) within-subject and either (ii) along the temporal dimension, or (iii) by local spatial smoothing. That is, for a fixed voxel (or set of voxels) we filter, smooth, detrend, despike, etc. [1]. In this paper we present a multi-subject, collaborative approach for denoising fMRI along the spatial domain. Rather than relying on local, isotropic spatial smoothing, our approach exploits shared between-subject temporal synchrony to denoise the data in a transformed spatial domain.

In [2], the authors observed that a common, synchronous stimulus, such as a movie, leads to increased inter-subject correlation (ISC). This suggests that we can move beyond *within-subject* methodologies and leverage *between-subject* commonality to jointly learn from each subject's fMRI reponse data. A successful example resulting from this line of reasoning is *hyperalignment*: a subject-specific, high-dimensional, orthogonal spatial mapping of fMRI data into a common abstract space [3]. The orthogonal constraints of hyperalignment (indirectly) lead to ISC maximization [4], and thus exploit shared between-subject structure. Within the hyperalignment common space, the performance of between-subject classification experiments matches or exceeds within-subject benchmarks, suggesting that the common space has extracted useful information shared across subjects.

Rather than use between-subject commonality for alignment purposes, here we explore whether denoising can also be accomplished. We begin by anatomically aligning the fMRI data across subjects. This alignment gives a common voxel coordinate system across subjects and a common voxel indexing and visualization mechanism. Moreover, a sparse set of voxels for one subject, defines the same sparse set of voxels for all subjects. Anatomical alignment has limitations, but it is often the default method for multi-subject analysis. We then ask whether the anatomically-aligned fMRI data can be denoised by exploiting hyperalignment. This leads to a collaborative denoising technique that uses part of the multi-subject fMRI data as training data to learn subject-specific orthogonal transformations. These transformations are then calibrated to align with the *principal axes* [5]. Whereas [5] used the data-induced principal axes as a visualization technique, similar to multidimensional scaling [6], we use these axes for signal denoising. Once in the transformed space, we apply established shrinkage methods along each dimension, which effectively denoises the data.

The remainder of the paper is organized as follows: we introduce our denoising method in §2 and offer a modeling approach of the problem, providing probabilistic insight in §3. §4 presents experimental results on multisubject fMRI data, and we conclude in §5.

2. PRINCIPAL AXES DENOISING

2.1. Hyperalignment

We begin with fMRI data collected for m subjects from a common, synchronous stimulus (e.g., viewing the *same* movie). The data are recorded in m matrices, $\mathbf{X}_{1:m} \in \mathbb{R}^{t\times n}$, where t is the number of TRs and n is the number of voxels. Note that the TRs index rows and the voxels index the columns. Given the synchronous stimulus, we assume row correspondence.

Hyperalignment constructs linear maps $\mathbf{R}_{1:m} \in \mathbb{R}^{n\times n}$ from voxel-space to a common space. This is accomplished by solving the multi-set orthogonal Procrustes problem:

$$\mathbf{R}_{1:m} = \underset{\mathbf{Q}_{1:m}\in\mathcal{O}(n)}{\arg\min} \sum_{i<j} \|\mathbf{X}_i\mathbf{Q}_i - \mathbf{X}_j\mathbf{Q}_j\|_{\mathrm{F}}^2 \ , \qquad (1)$$

where $\mathcal{O}(n)$ is the set of $n \times n$ orthogonal matrices. Orthog-

onality ensures temporal isometry, i.e.,

$$\mathbf{X}_i\mathbf{X}_i^T \mapsto (\mathbf{X}_i\mathbf{R}_i)(\mathbf{X}_i\mathbf{R}_i)^T = \mathbf{X}_i\mathbf{X}_i^T.$$

The implication is that the temporal geometry associated with the stimulus is invariant within a subject. For solving (1), the reader is referred to [7, 8, 9].

2.2. Learning a basis

In traditional denoising, we project the data $\mathbf{x}$ onto a predefined orthonormal (ON) basis, yielding coordinates $\hat{\mathbf{x}}$. The basis is often specifically chosen to yield a sparse data representation, i.e., many entries of $\hat{\mathbf{x}}$ are zero (or nearly zero). For example, to denoise audio we might project onto a Fourier basis and for images we might project onto a wavelet basis. Motivated by the assumed sparsity of $\hat{\mathbf{x}}$, after mapping into the transform domain we denoise by shrinking the coefficients $\hat{\mathbf{x}}$ toward zero. We then apply the inverse transform to obtain a denoised signal.

Under hyperalignment, since $\mathbf{R}_i$ is orthogonal an ON basis is being learnt for each subject. This learning process attempts to match subjects' fMRI responses through their subject-specific transform representation. That is, under the hypothesis that $\mathbf{X}_i\mathbf{R}_i \approx \mathbf{X}_j\mathbf{R}_j$, the optimization problem of (1) seeks the $\mathbf{R}_{1:m}$ that minimizes the sum of all pairwise distances $\|\mathbf{X}_i\mathbf{R}_i - \mathbf{X}_j\mathbf{R}_j\|_{\mathrm{F}}^2$.

Once $\mathbf{R}_{1:m}$ is evaluated we may be tempted to denoise subject i via $\mathbf{R}_i$ and the process described above. However, there is no justification for assuming a sparse representation of subject i's data after projecting onto the ON basis vectors in $\mathbf{R}_i$. We also want sparsity in the transformed domain.

The unitary invariance of the Frobenius norm means that there is a family of solutions for (1). If $\mathbf{R}_{1:m}$ is a global solution then so is $\{\mathbf{R}_1\mathbf{R}_0, \ldots, \mathbf{R}_m\mathbf{R}_0\}$ for *any* $\mathbf{R}_0 \in \mathcal{O}(n)$. Consider a solution $\{\mathbf{R}_1\mathbf{R}_0, \ldots, \mathbf{R}_m\mathbf{R}_0\}$. We can achieve sparsity by using $\mathbf{R}_0$ to seek a sparse centroid or *consensus matrix* $\mathbf{Y} = \frac{1}{m}\sum_{i=1}^m \mathbf{X}_i\mathbf{R}_i\mathbf{R}_0$. The following identity is the key to this goal [7]:

$$\sum_{i<j} \|\mathbf{X}_i\mathbf{R}_i\mathbf{R}_0 - \mathbf{X}_j\mathbf{R}_j\mathbf{R}_0\|_{\mathrm{F}}^2 = m\sum_{i=1}^m \|\mathbf{X}_i\mathbf{R}_i\mathbf{R}_0 - \mathbf{Y}\|_{\mathrm{F}}^2. \quad (2)$$

So instead of considering pairwise matchings, we consider how the transformed data relates to the centroid ("total distance versus variance" [10]). Suppose $\mathbf{Y} \in \mathbb{R}^{t\times n}$ has rank $p \le \min(t, n)$—a property independent of $\mathbf{R}_0$. We can use $\mathbf{R}_0$ to ensure that the p left singular vectors of $\mathbf{Y}$ are positioned in the first p columns of $\mathbf{Y}$. Specifically, let $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{I}^T = \mathbf{U}\mathbf{\Sigma}$ be the *full* SVD of $\mathbf{Y}$ with $\mathbf{U} \in \mathbb{R}^{t\times p}$ housing the p left singular vectors. This provides the sparsest representation of the centroid in terms of orthogonal vectors. The left singular vectors are also referred to as the *principal axes*.

To achieve the desired form for $\mathbf{Y}$, we start from the original centroid $\bar{\mathbf{Y}} = \frac{1}{m}\sum_{i=1}^m \mathbf{X}_i\mathbf{R}_i$. Let the *full* SVD of $\bar{\mathbf{Y}}$ be $\bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\bar{\mathbf{V}}^T$. If we set $\mathbf{R}_0 = \bar{\mathbf{V}} \in \mathcal{O}(n)$, we obtain

$$\mathbf{Y} = \frac{1}{m}\sum_{i=1}^m \mathbf{X}_i\mathbf{R}_i\mathbf{R}_0 = \left(\frac{1}{m}\sum_{i=1}^m \mathbf{X}_i\mathbf{R}_i\right)\mathbf{R}_0 \quad (3)$$

$$= \bar{\mathbf{Y}}\mathbf{R}_0 = \left(\bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\bar{\mathbf{V}}^T\right)\bar{\mathbf{V}} = \bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\ , \quad (4)$$

leading to the desired form described above.

Algorithm 1 Basis Discovery

Input: $\mathbf{X}_{1:m} \in \mathbb{R}^{t\times n}$
Output: ON bases $\mathbf{R}_{1:m}$
$\tilde{\mathbf{X}}_i \overset{1:m}{\longleftarrow} \mathbf{X}_i\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)$
$\mathbf{R}_{1:m} \longleftarrow \arg\min_{\mathbf{Q}_{1:m}\in\mathcal{O}(n)} \sum_{i<j}\|\tilde{\mathbf{X}}_i\mathbf{Q}_i - \tilde{\mathbf{X}}_j\mathbf{Q}_j\|_{\mathrm{F}}^2$
$\bar{\mathbf{Y}} \longleftarrow \frac{1}{m}\sum_{i=1}^m \tilde{\mathbf{X}}_i\mathbf{R}_i$
$[\bar{\mathbf{U}}\ \bar{\mathbf{\Sigma}}\ \bar{\mathbf{V}}] \longleftarrow \mathrm{SVD}(\bar{\mathbf{Y}})$
$\mathbf{R}_i \overset{1:m}{\longleftarrow} \mathbf{R}_i\bar{\mathbf{V}}$

Before proceeding to the denoising aspect we address mean invariance. When we think of Fourier denoising, we typically focus on the AC components (frequency > 0) and preserve the DC component (frequency $= 0$). We can apply the same methodology here as well. Before solving (1) we first subtract the temporal mean, $\frac{1}{n}\mathbf{X}_i\mathbf{1}$, from each column of $\mathbf{X}_i$ ($i = 1, \ldots, n$). Subsequent operations will transform the data in a space orthogonal to the averaging vector $\mathbf{1}$. Algorithm 1 presents the basis discovery for $\mathbf{X}_{1:m}$. Here $\tilde{\mathbf{X}}_i$ is $\mathbf{X}_i$ with the mean projected out.

2.3. Denoising

Algorithm 1 learns a basis from training data $\mathbf{X}_{1:m} \in \mathbb{R}^{t\times n}$. The next step is to denoise new data $\mathbf{Z}_{1:m} \in \mathbb{R}^{s\times n}$ —both $\mathbf{X}_i$ and $\mathbf{Z}_i$ are data from the i-th subject possessing column correspondence, i.e. within-subject alignment. Adhering to mean invariance, we focus on a filter for $\tilde{\mathbf{Z}}_i = \mathbf{Z}_i\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)$.

If we use a linear filter equipped with predefined weights $\mathbf{c} \in \mathbb{R}^n$, then the denoised signal is $\tilde{\mathbf{Z}}_i\mathbf{R}_i\,\mathbf{diag}\{\mathbf{c}\}\mathbf{R}_i^T$. This is composed of the following steps: we first map the signal into the transform domain ($= \tilde{\mathbf{Z}}_i\mathbf{R}_i$), multiply the k-th column by c_k ($= \tilde{\mathbf{Z}}_i\mathbf{R}_i\,\mathbf{diag}\{\mathbf{c}\}$), and then apply the inverse transform ($= \tilde{\mathbf{Z}}_i\mathbf{R}_i\,\mathbf{diag}\{\mathbf{c}\}\mathbf{R}_i^T$).

Typically, we select each $c_k \in [0, 1]$; when $c_k = 1$ there is no denoising along the k-th principal axis, and when $c_k = 0$ there is full denoising along the k-th principal axis. The Fourier basis is naturally *ordered* by the notion of frequency, so that c_k can be set to yield a low-pass filter, for example, by setting $c_k = 1$ when $k < k_0$ and 0 otherwise. In contrast, the ON bases $\mathbf{R}_{1:m}$ are not assumed to possess any type of ordering. Wavelet representations have a similar property in that, within a scale, any ordering is valid. Thus, for $\mathbf{W}_i = \tilde{\mathbf{Z}}_i\mathbf{R}_i = \left[\mathbf{w}_i^1 \mid \mathbf{w}_i^2 \mid \cdots \mid \mathbf{w}_i^n\right]$, we consider adaptive

Algorithm 2 Basis Denoising

1: **Input:** ON bases $\mathbf{R}_{1:m} \in \mathcal{O}(n)$; noisy data $\mathbf{Z}_{1:m} \in \mathbb{R}^{s \times n}$
2: **Output:** denoised data $\hat{\mathbf{Z}}_{1:m}$
3: $\mathbf{b}_i \overset{1:m}{\longleftarrow} \frac{1}{n}\mathbf{Z}_i\mathbf{1}$
4: $\tilde{\mathbf{Z}}_i \overset{1:m}{\longleftarrow} \mathbf{Z}_i - \mathbf{b}_i\mathbf{1}^T$
5: $\mathbf{W}_i \overset{1:m}{\longleftarrow} \tilde{\mathbf{Z}}_i\mathbf{R}_i$
6: (note: $\mathbf{W}_i = \left[\mathbf{w}_i^1 \mid \mathbf{w}_i^2 \mid \cdots \mid \mathbf{w}_i^n\right]$)
7: **for** $i = 1 \to m$ **do**
8: **for** $j = 1 \to n$ **do**
9: $\mathbf{w}_i^j \longleftarrow$ shrink($\mathbf{w}_i^j$)
10: **end for**
11: **end for**
12: $\tilde{\mathbf{Z}}_i \overset{1:m}{\longleftarrow} \mathbf{W}_i\mathbf{R}_i^T$
13: $\hat{\mathbf{Z}}_i \overset{1:m}{\longleftarrow} \tilde{\mathbf{Z}}_i + \mathbf{b}_i\mathbf{1}$

coefficients of the form [11]:

$$c_k = \max\left\{1 - \left(\frac{\tau}{\|\mathbf{w}_i^k\|}\right)^{\beta}, 0\right\}, \quad (5)$$

where $\tau \geq 0$ is a threshold parameter and $\beta > 0$. Familiar forms include hard thresholding ($\beta \to +\infty$), soft thresholding ($\beta = 1$), and James-Stein shrinkage or empirical Wiener attenuation ($\beta = 2$). Algorithm 2 features the denoising scheme. For flexibility, line 9 is left generic; shrinkage can be accomplished with constant or adaptive weights.

3. A MODELING PERSPECTIVE

In [12], the authors revealed the emergence of two function brain networks: extrinsic and intrinsic. The extrinsic network is tied to the stimulus while the intrinsic network resembles the default mode network [13]. Thus, under a common stimulus extrinsic networks may exhibit common structure across subjects, while the intrinsic networks remain isolated.

Suppose $\mathbf{X}_0 \in \mathbb{R}^{t \times n}$ is the noiseless, universal response for a given stimulus and model the recorded responses as $\mathbf{X}_i = \mathbf{X}_0\mathbf{R}_i^T + \mathbf{H}_i$, where $\mathbf{H}_i$ is a noise matrix that may also include the effect of the intrinsic network. In the spirit of hyperalignment, the $\mathbf{R}_i$ denote isometric transformations to account for alignment differences.

We use the " $\breve{}$ " accent to denote random variables. For two datasets i and j, we make the assumption

$$\mathbb{E}\{\breve{h}_i(k,l)\} = 0 \quad \text{and} \quad \mathbb{E}\{\breve{h}_i(k,l)\breve{h}_j(k',l')\} = 0\,, \quad (6)$$

where $h_i(k,l)$ is the kl-th entry of matrix $\mathbf{H}_i$. It follows that

$$\mathbb{E}\{\breve{\mathbf{X}}_i\} = \mathbb{E}\{\mathbf{X}_0\mathbf{R}_i^T + \breve{\mathbf{H}}_i\} = \mathbf{X}_0\mathbf{R}_i^T \quad (7)$$

$$\begin{aligned}\mathbb{E}\{\breve{\mathbf{X}}_i^T\breve{\mathbf{X}}_j\} &= \mathbb{E}\{(\mathbf{X}_0\mathbf{R}_i^T + \breve{\mathbf{H}}_i)^T(\mathbf{X}_0\mathbf{R}_j^T + \breve{\mathbf{H}}_j)\} \\ &= \mathbf{R}_i\mathbf{X}_0^T\mathbf{X}_0\mathbf{R}_j^T\,. \quad (8)\end{aligned}$$

Note, however, that $\mathbb{E}\{\breve{\mathbf{X}}_i^T\breve{\mathbf{X}}_i\} = \mathbf{R}_i\mathbf{X}_0^T\mathbf{X}_0\mathbf{R}_i^T + \mathbb{E}\{\breve{\mathbf{H}}_i^T\breve{\mathbf{H}}_i\}$, with $\mathbb{E}\{\breve{\mathbf{H}}_i^T\breve{\mathbf{H}}_i\}$ a quantity we assume no prior knowledge about. Even when all of the latent $\mathbf{R}_{1:m}$ are identity matrices, the characteristics of $\breve{h}_i(k,l)$ are still unknown.

Under the orthogonal constraint, it can be shown [4] that (1) is equivalent to:

$$\mathbf{R}_{1:m} = \underset{\mathbf{Q}_{1:m} \in \mathcal{O}(n)}{\arg\max} \textstyle\sum_{i<j} \frac{1}{n} \operatorname{tr}\left(\mathbf{R}_i^T\mathbf{X}_i^T\mathbf{X}_j\mathbf{R}_j\right)\,. \quad (9)$$

Since the ISC between subject i and j is $\frac{1}{n}\operatorname{tr}(\mathbf{X}_i^T\mathbf{X}_j)$, the above relationship indicates that hyperalignment is maximizing the ISC in the common space via the transformation $\mathbf{X}_i^T\mathbf{X}_j \mapsto (\mathbf{X}_i\mathbf{R}_i)^T\mathbf{X}_j\mathbf{R}_j$. Additionally, in the form provided in (9), we observe that $\mathbf{X}_i^T\mathbf{X}_i$ is absent—the objective does not consider intra-subject connectivity, and therefore may be immune to intrinsic network effects [12]. Furthermore, we have $\mathbb{E}\{\mathbf{R}_i^T\breve{\mathbf{X}}_i^T\breve{\mathbf{X}}_j\mathbf{R}_j\} = \mathbf{X}_0^T\mathbf{X}_0$, so the above objective is not biased by the noise.

4. EXPERIMENTS

Talairach-aligned fMRI data from ten subjects were collected as they watched the movie *Raiders of the Lost Ark* (1981) and partook in a visual block-design experiment [14]. The block-design experiment featured 7 categories of images: (1) female-face, (2) male-face, (3) monkey-face, (4) dog-face, (5) house, (6) chair, and (7) shoe. In each run, a subject was shown all 7 image categories, with each category featured for 16 TRs followed by 10 TRs of rest (1 TR = 2 seconds). There were 8 runs per subject.

In our experiments we focus on Ventral Temporal (VT) cortex. We took the time average of all 16 TR windows where images were displayed (offset by 2 TRs to account for hemodynamic response). This produced a labeled dataset of 560 examples (10 subjects $\cdot$ 8 runs/subject $\cdot$ 7 examples/run) with 5994 features (2997 voxels per hemisphere of VT).

We investigated whether we can denoise the second half of the movie by learning ON bases from the first half. In particular, we look at the average ISC as we move from no denoising to complete denoising. Recall that no denoising means we leave the data untouched, while complete denoising means that, for a given TR, each voxel is replaced by the average response.

The ISC is a between-subject metric. Playing an adversarial role, if we "denoise" each voxel time-series by replacing it with same fixed signal, the average ISC will be 1 and therefore lead to erroneous conclusions about our denoising. Thus, we also need a within-subject metric to ensure that we have not lost signal information. To accomplish this, we also computed the cross-validated within-subject classification (WSC) accuracy on the 7-label dataset, using the same per-subject filter coefficients used for the movie denoising (for classification we used a one-versus-one ν-SVM [15]). In this way,

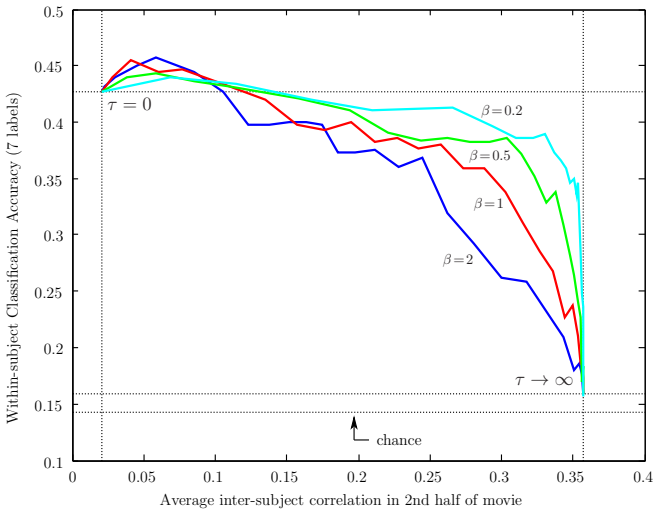

Fig. 1. Denoising: ISC vs WSC for 7-label data

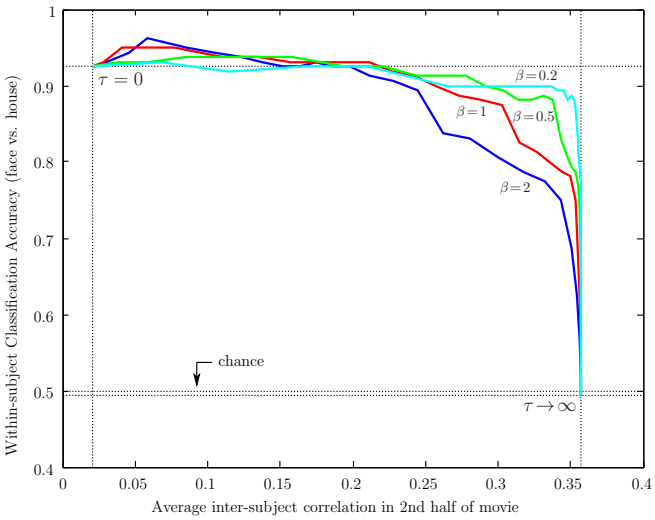

Fig. 2. Denoising: ISC vs WSC for face-vs-house

if a large ISC is accompanied by chance within-subject accuracy (=1/7), we know that the denoising has removed crucial information from the original signal and cannot be trusted.

Figures 1 and 2 feature the improvement in ISC using the shrinkage of (5) for varying values of β. Each path is parametrized by τ as it increases from 0 to $+\infty$ (effectively). At $\tau = 0$, there is no denoising as the filter coefficients are all equal to 1. Similarly, when $\tau \to +\infty$ all of the filter coefficients are 0 and only the mean response survives. Thus, all paths have the same starting (ending) points, irrespective of β. Figure 1 keeps us honest by tracking the WSC accuracy of the 7-label dataset; Figure 2 considers WSC accuracy of the classification of face versus house.

When $\beta = 0.2$ the mean WSC accuracy for the 7-label dataset is maintained while increasing the mean ISC from 0.02 ± 0.002 to 0.14 ± 0.01 —a 600% increase. For face vs house WSC, the mean ISC increases to 0.21 ± 0.01 without loss of discriminative information. Starting at $\tau = 0$, the WSC accuracy increases and then decreases. The increase indicates that denoising is indeed improving the signal-to-noise ratio. There is also a tradeoff between β and "dynamic range" of ISC. For β large there is a larger peak in WSC accuracy followed by relatively steep descent. For β small, the WSC remains high for a longer range of ISC, but WSC accuracy does not increase as much.

Alignment	(β, τ)	WSC Acc.	BSC Acc.
Talairach	—	42.68%	35.71%
	(1.0, 0.9)	44.82%	36.07%
Hyperalignment	—	42.68%	45.71%
	(1.0, 0.9)	44.82%	48.93%

Table 1. Denoising on Talairach and hyperaligned data and 7-label classification. Standard errors $\approx 2\%$.

Next, we looked at a specific value of (β, τ) and analysed the effects of denoising for both Talairach and hyperalignment. Although not discussed, denoising in hyperaligned space actually involves fewer steps because we never leave the common space. Table 1 provides WSC and Between Subject Classification (BSC) accuracy for the 7-label classification before and after denoising. With $(\beta, \tau) = (1.0, 0.9)$ accuracy improved, albeit by a small amount, for both alignment spaces. This is significant for between subject classification under hyperalignment.

5. CONCLUSION

Traditional spatial averaging uses local isotropic filters to spatially smooth each subject's data. This improves the local signal-to-noise ratio for each subject separately at the expense of spatial resolution. In contrast, here we exploited the across-subject shared component of fMRI responses elicited by a common (synchronous) stimulus. This was identified via a hyperalignment training phase and then used to appropriately pool the subjects' data to permit collaborative denoising. The resultant algorithms execute quickly, and on real multi-subject fMRI data were able to improve inter-subject correlation without reducing within subject classification accuracy—an indication that noise was being removed without detriment to the information present.

6. REFERENCES

[1] R.W. Cox, "AFNI: software for analysis and visualization of functional magnetic resonance neuroimages,"

Computers and Biomedical Research, vol. 29, no. 3, pp. 162–173, 1996.

[2] U. Hasson, Y. Nir, I. Levy, G. Fuhrmann, and R. Malach, "Intersubject synchronization of cortical activity during natural vision," *Science*, vol. 303, no. 5664, pp. 1634–1640, 2004.

[3] J.V. Haxby, J.S. Guntupalli, A.C. Connolly, Y.O. Halchenko, B.R. Conroy, M.I. Gobbini, M. Hanke, and P.J. Ramadge, "A common, high-dimensional model of the representational space in human ventral temporal cortex," *Neuron*, vol. 72, no. 2, pp. 404–416, 2011.

[4] H. Xu, A. Lorbert, P.J. Ramadge, J.S. Guntupalli, and J.V. Haxby, "Regularized hyperalignment of multi-set fMRI data," in *IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2012, pp. 229–232.

[5] G.M. Arnold and A.J. Collins, "Interpretation of transformed axes in multivariate analysis," *Appl. Stat.*, vol. 42, no. 2, pp. 381–400, 1993.

[6] J.B. Tenenbaum, V. De Silva, and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[7] J.C. Gower and G.B. Dijksterhuis, *Procrustes Problems*, vol. 30, Oxford University Press, 2004.

[8] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2009.

[9] A. Lorbert and P.J. Ramadge, "Kernel hyperalignment," in *Advances in Neural Information Processing Systems 25*, P. Bartlett, F.C.N. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, Eds., pp. 1799–1807. 2012.

[10] D.R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, 2004.

[11] S. Mallat, *A Wavelet Tour of Signal Processing, The Sparse Way*, Elsevier Science, 2009.

[12] Y. Golland, S. Bentin, H. Gelbard, Y. Benjamini, R. Heller, Y. Nir, U. Hasson, and R. Malach, "Extrinsic and intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation," *Cerebral Cortex*, vol. 17, no. 4, pp. 766–777, 2007.

[13] D.A. Gusnard and M.E. Raichle, "Searching for a baseline: functional imaging and the resting human brain," *Nature Reviews Neuroscience*, vol. 2, no. 10, pp. 685–694, 2001.

[14] M.R. Sabuncu, B.D. Singer, B. Conroy, R.E. Bryan, P.J. Ramadge, and J.V. Haxby, "Function-based intersubject alignment of human cortical anatomy," *Cerebral Cortex*, vol. 20, no. 1, pp. 130–140, 2010.

[15] P.H. Chen, C.J. Lin, and B. Schölkopf, "A tutorial on ν-support vector machines," *Appl. Stoch. Models Bus. Ind.*, vol. 21, no. 2, pp. 111–136, 2005.