Bayesian source separation of fMRI signals

Daniel B. Rowe

California Institute of Technology
MSC 228-77
1200 E. California Blvd.
Pasadena, California 91125

Abstract. In analyzing the results of functional magnetic resonance imaging, the identification of significant activation in voxels is a crucial task. In computing the activation level, a standard method is to select an assumed to be known reference function and perform a multiple regression of the time courses on it and a linear trend. Once the linear trend is found, the correlation between the assumed to be known reference function and the detrended observed time-course in each voxel is computed and voxels colored according to their correlation. But the most important question is: How do we choose the reference function? This paper develops a Bayesian statistical approch to determining the underlying source reference function based on Bayesian source separation, and uses it on both sumulated and real fMRI data. This underlying reference function is the unobserved response due the presentation of the experimental stimulus.

INTRODUCTION

Typically in an fMRI, the patient is given a sequence of two stimuli or tasks, A and B. Imaging is taking place while the patient is responding either passively or actively to these tasks. A linear model is used to describe the observed signal in each voxel. The model views the observed signal as being made up of a linear trend, a response (possibly zero valued) due to the tasks, and other cognitive activity that is typically termed random and grouped into the error term. The association between the observed time course in each voxel and the sequence of tasks is determined. Different levels of activation are assigned coloration and the voxels colored according to their activation.

In computing the activation level in a given voxel, a standard method [1, 2] is to compute the correlation between an assumed to be known reference function (typically a square, triangle, or sine wave with the same period as the experimental sequence) and the detrended observed time-course in each voxel. But the most important question is: How do we choose the reference function? The reference function is viewed as the underlying response due to the presentation of the experimental tasks. What if the response does not fit into the standard on/off or rise/fall format? What if it changes (possibly nonlinearly) over the course of the experiment? The choice of the reference function in computing the activation in fMRI is somewhat arbitrary and subjective. This paper develops a coherent Bayesian statistical approch to determine the underlying response or reference function. In this approach, all the voxels contribute to "telling us" the underlying response due to the experimental stimulus.

CP568, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 20th International Workshop,* edited by A. Mohammad-Djafari

To motivate the source separation model, consider the context of the classic "cocktail party" problem [3, 4]. At a cocktail party, there are microphones scattered about that record partygoers or speakers at a given number of time increments. The observed conversations consist of mixtures of true unobservable conversations. In other words, at each time increment mixed signal vectors are observed and the goal is to separate these observed signal vectors into true unobservable underlying source signal vectors. This is exactly the problem we are addressing in fMRI. The source separation model decomposes the observed time course in a voxel into a linear trend and a linear combination of unobserved component sequences. If there was only one response function or component time sequence and it is assumed to be known, then the Bayesian approach reduces to the standard model and the correlation technique may be implemented. In practice we do not know the true underlying time response function.

The Bayesian source separation model assesses a prior mean for the response function, combines it with the data, and computes a posterior mean response. The correlation technique may now be implemented between the posterior mean response and the detrended time sequence in each voxel. The Bayesian source separation model [5] also allows for several different and possibly correlated component time sequences. There are always incidental cognitive processes and blood flow that may be considered as components. These time sequences could correspond to activity such as that due to an EKG and respiration. Modeling them instead of grouping them into the error term could prove useful.

MODEL

Consider the observed value in voxel j at time i, the model is

$$x_{ij} = a_j + b_j i + \lambda_{j1} s_{i1} + \ldots + \lambda_{jm} s_{im} + \varepsilon_{ij}. \tag{1}$$

That is, the observed signal in voxel j at time i contains a linear part with a slope and intercept in addition to a linear combination of the m unobserved source components $s_{i1}, \ldots, s_{im}$ with amplitudes or mixing coefficients $\lambda_{j1}, \ldots, \lambda_{jm}$. This model can be written in terms of vectors as

$$x_{ij} = \beta_j' u_i + \lambda_j' s_i + \varepsilon_{ij}. \tag{2}$$

where $u_i = (1, i)'$, $\beta_j = (a_j, b_j)'$, $\lambda_j = (\lambda_{j1}, \ldots, \lambda_{jm})'$ and $s_i = (s_{i1}, \ldots, s_{im})'$. If any or all of the sources were assumed to be known, they could be grouped into the u's and their coefficients computed.

Each voxel has its own slope and intercept in addition to a set of mixing coefficients that do not change over time. In contrast, the unobserved underlying source reference functions are the same for all voxels (with possibly zero valued coefficients) at a given time but do change over time.

Now, considering all p voxels at time i, the model can be written as

$$x_i = \beta u_i + \Lambda s_i + \varepsilon_i \tag{3}$$

where $B = (\beta_1, \ldots, \beta_p)'$, is an $p \times 2$ matrix of slopes and intercepts, and $\Lambda = (\lambda_1, \ldots, \lambda_p)'$ is an $p \times m$ dimensional matrix of mixing coefficients.

The model which considers all the voxels at all time points can be written in terms of matrices as

$$X = UB' + S\Lambda' + E \tag{4}$$

where $X' = (x_1, \ldots, x_n)$, $U = (e_n, c_n)$, e_p is a $p \times 1$ vector of ones, $c_n = (1, \ldots, n)'$, $S' = (s_1, \ldots, s_n)$, $E' = (\varepsilon_1, \ldots, \varepsilon_n)$ while B ans Λ are as before.

Motivated by the central limit theorem, the errors of observation are taken to be normally distributed, as $(\varepsilon_i|\Psi) \sim N(0, \Psi)$, then the observations are also normally distributed as

$$p(X|B, S, \Lambda, \Psi) \propto |\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2} tr \Psi^{-1} (X - UB' - S\Lambda')'(X - UB' - S\Lambda')} \tag{5}$$

PRIORS

Recall that the source components, the s_i's are unobserved. As previously noted, the typical method for determining activations in each voxel is to subjectively assign one source reference function. This reference function is commonly chosen to be either a square, triangle, or sine wave and sometimes shifted. Once the reference function is chosen, a regression is performed to fit the model

$$x_{ij} = a_j + b_j i + \lambda_j s_i + \varepsilon_{ij} \tag{1}$$

and obtain the regression coefficient estimates $(\hat{a}_j, \hat{b}_j, \hat{\lambda}_j)$. Significant activation is determined by correlation of the reference function with the voxels detrended time course and voxels are assigned coloration according to their activation level.

The above method of subjectively asigning a source reference function and performing regression is equivalent to assigning a degenerate distribution for it. That is, equivalent to assuming that the probability distribution for the source reference function is equal unity at this assigned value and zero otherwise.

Instead of subjectively choosing a source reference function, prior information as to its value in the form of a prior distribution is assessed (as are priors for any other contributing source reference functions to the observed signal). This prior distribution is combined with the data and a source reference function is determined statistically using the information contributed from every voxel. In addition, prior distributions are assessed for the the covariance matrix for the sources, the slopes and intercepts, the mixing coefficients, and the covariance matrix for the observation error.

When quantifing available prior information regarding the parameters of interest, natural conjugate prior distributions [**? ?**] are specified. The prior distribution for the source reference functions is taken to be a normal distribution where the source components are uncorrelated over time but correlated at a given time. The mixing coefficients are also taken to be normally distributed, while the observation and source covariance matrices are taken to be Inverse Wishart distributed. A vague or noninformative distribution is taken for the regression coefficients.

The prior distributions for the parameters of interest are

$$p(S|R) \propto |R|^{-\frac{n}{2}} e^{-\frac{1}{2}trR^{-1}(S-S_0)'(S-S_0)} \quad (2)$$

$$p(R) \propto |R|^{-\frac{\eta}{2}} e^{-\frac{1}{2}trR^{-1}V} \quad (3)$$

$$p(\Lambda|\Psi) \propto |A|^{-\frac{p}{2}} |\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\Lambda_0)A^{-1}(\Lambda-\Lambda_0)'} \quad (4)$$

$$p(\Psi) \propto |\Psi|^{-\frac{\nu}{2}} e^{-\frac{1}{2}tr\Psi^{-1}Q}, \quad (5)$$

$$p(B) \propto \text{constant} \quad (6)$$

where the prior mean for the source component reference functions is $S_0' = (s_{01}, \ldots, s_{0n})$. The prior distributions for the slopes and intercepts could easily be taken to be normal.

POSTERIOR AND ESTIMATION

Note that the data consists of 98304 time courses each of length 128. This data will be combined with the prior distributions to produce a joint posterior distribution. With this sample size, the prior distributions will have little influence in the final results. The posterior distribution is robust with respect to prior specifications. Using Bayes' theorem, the posterior distribution can be written asbeing proportional to the aformentioned priors and likelihood, but it can not be integrated analytically to obtain marginal densities and thus marginal estimates. However, maximum a posteriori estimates can be obtained via the iterated conditional modes (ICM) algorithm [6, 7, 8]. This algorithm uses the posterior conditional distributions and cycles through their modes.

Maximization of the full posterior distribution with respect to each of the parameters is equivalent to maximization of each of the posterior distributions. The ICM algorithm consists of starting with an initial S and B values, say $S_{(0)}$ and $B_{(0)}$ then iterate through

$$\tilde{\Lambda}_{(l+1)} = [\Lambda_0 A^{-1} - (X - UB'_{(l)})'S_{(l)}](A^{-1} + S'_{(l)}S_{(l)})^{-1} \quad (1)$$

$$\tilde{\Psi}_{(l+1)} = \frac{1}{n+m+\nu}\Big\{(X - UB'_{(l)} - S_{(l)}\Lambda'_{(l+1)})'(X - UB'_{(l)} - S\Lambda'_{(l+1)}) + (\Lambda_{(l+1)} - \Lambda_0)A^{-1}(\Lambda_{(l+1)} - \Lambda_0)' + Q\Big\} \quad (2)$$

$$\tilde{R}_{(l+1)} = \frac{(S_{(l)} - S_0)'(S_{(l)} - S_0) + V}{n+\eta} \quad (3)$$

$$\tilde{S}_{(l+1)} = [S_0 R^{-1}_{(l+1)} + (X - UB'_{(l)})\Psi^{-1}_{(l+1)}\Lambda_{(l+1)}](R^{-1}_{(l+1)} + \Lambda'_{(l+1)}\Psi^{-1}_{(l+1)}\Lambda_{(l+1)})^{-1} \quad (4)$$

$$\tilde{B}_{(l+1)} = (X - S_{(l+1)}\Lambda'_{(l+1)})'U(U'U)^{-1} \quad (5)$$

until convergence is reached which may be defined to be until each changes by less than any desired precision from the previous iteration.

As noted before, a typical number of voxels is $p = 64\text{x}64\text{x}24 = 98304$. With this many voxels, Ψ has $\frac{p(p+1)}{2} = 4.8\text{x}10^9$ distinct covariance elements. This is too big for practical purposes due to the fact that in the ICM algorithm, Ψ must be inverted. For

this reason, it is currrently assumed cor computational purposes only that the voxels are spatially uncorrelated. This is an assumption for computational purposes and not a model assumption. Relaxation of the assumption of uncorrelated voxels for practical purposes is being pursued. Local correlation is being investigated as are sparce matrix inversion techniques.

After determining the Bayesian reference function associated with the experimental task, the correlation coefficient is calculated with each voxels detrended observed time course. A threshold is set and if the correlation is below the threshold, then its value is set to zero. If it is above the threshold, then its value is retained. For values above the threshold, a one to one color mapping is performed. The image of the colored voxels is superimposed onto an anatomical image.

SIMULATED FMRI EXAMPLE

For an example, data was generated to mimic a scaled down fMRI experiment. The simulated experiment is chosen to have two tasks each of length 32 seconds with eight rounds totaling 512 seconds as illustrated in Figure 1.

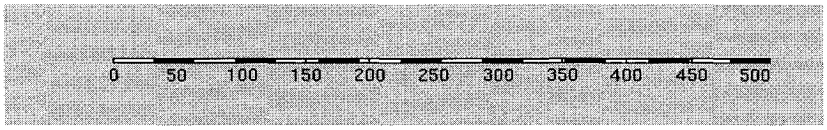

FIGURE 1. Experimental Design: White Task A and Black Task B in seconds.

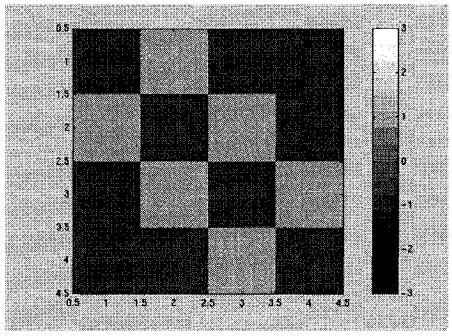

FIGURE 2. Anatomical Image.

An anatomical 4x4 image is determined as in Figure 2. As in a typical fMRI experiment, observations will be taken every four seconds so that there are 128 in each voxel. The functional data is created with a known source reference function. This reference function is a unit amplitude sinusoid with a frequency of $1/64$ Hz.

In addition, a second source with unit amplitude and frequency $80/60$ Hz was added to corrupt the signal as was random noise for each. The voxels are numbered from one to

sixteen starting from the top left and proceeding across then down. The functional data for these 16 voxels was created according to the source separation model

$$\underset{(1\times 1)}{x_{ij}} = \underset{(1\times 2)}{\beta'_{jT}}\ \underset{(2\times 1)}{u_i} + \underset{(1\times 2)}{\lambda'_j}\ \underset{(2\times 1)}{s_i} + \underset{(1\times 1)}{\varepsilon_{ij}} \tag{1}$$

where j denotes the voxel, i denotes the time increment. and ε_{ij} denotes the random error term. The random errors were generated according to $\varepsilon_{ij} \sim N(0, 10)$. Noise was added to the sources reference functions and the mixing coefficients according to $s_i \sim N(s_{Ti}, \frac{1}{25} I_2)$ and $\lambda_j \sim N(\lambda_{Tj}, \frac{1}{16} I_2)$.

The true slopes and intercepts for the voxels along with the true amplitudes are displayed in their voxels location as

$$B_T : \begin{pmatrix} .2,.5 & .7,.1 & .4,.9 & .3,.2 \\ .9,.6 & .4,.8 & .5,.3 & .2,.7 \\ .9,.1 & .1,.3 & .5,.5 & .1,.6 \\ .6,.4 & .4,.2 & .4,.5 & .8,.9 \end{pmatrix}, \text{and } \Lambda_T : \begin{pmatrix} 15,1 & 2,2 & 1,1 & 2,2 \\ 1,1 & 15,2 & -5,1 & 2,2 \\ 1,1 & -5,2 & 15,1 & 2,1 \\ 1,2 & 2,2 & 1,1 & 15,2 \end{pmatrix}.$$

It is assumed that there is only one source reference function. In generating the data, the second one was added only as noise. All hyperparameters were assessed according to the empirical Bayes regression approach in Appendix A except for the prior mean and variance for the reference function. For the prior mean, a square wave is assessed with unit amplitude and frequency $1/64$ Hz which mimics the experiment.

For presentation purposes, all values have been rounded to two digits. The assessed prior intercepts and slopes were

$$B_0 : \begin{pmatrix} -1.68,0.50 & 1.09,0.08 & -0.92,0.93 & -0.48,0.23 \\ 1.87,0.58 & 1.03,0.80 & 0.32,0.30 & 0.98,0.68 \\ 1.96,0.09 & 0.88,0.30 & 1.36,0.47 & 1.89,0.57 \\ 1.71,0.40 & -1.14,0.22 & -0.41,0.51 & 1.70,0.89 \end{pmatrix}$$

The assessed prior mixing coefficients from were

$$\Lambda_0 : \begin{pmatrix} 10.35 & 2.09 & 0.51 & 0.96 \\ 0.08 & 9.02 & -3.88 & 0.80 \\ -0.28 & -2.73 & 10.32 & 1.33 \\ 0.62 & 0.61 & 2.28 & 9.70 \end{pmatrix}.$$

The assessed prior values for the mean and variance of the error are $\hat{E}(\psi^2) = 106.61$ and $\widehat{var}(\psi^2) = 490.64$ so that $\nu = 6.11$ and $q_0 = 224.79$. The prior mean and variance for the reference function was subjectively assessed to be $\hat{E}(r_{kk}) = (\frac{3}{4})^2$ and $\widehat{var}(r_{kk}) = 100$ which defined $\eta = 6.01$ and $v_0 = 10.13$.

Using these hyperparameter specifications, the Bayesian maximum a posteriori slopes and intercepts were found to be

$$\tilde{B} : \begin{pmatrix} -2.09,0.50 & 1.06,0.08 & -1.03,0.93 & -0.48,0.23 \\ 1.76,0.58 & 0.46,0.81 & 0.52,0.30 & 0.86,0.68 \\ 1.96,0.09 & 0.93,0.30 & 0.83,0.48 & 1.78,0.58 \\ 1.71,0.40 & -1.06,0.22 & -0.43,0.51 & 1.23,0.90 \end{pmatrix}.$$

As a measure of comparison of the different methods, the MSE's were computed. The MSE between the true linear coeficients and those from the standard regression was $mse(B_T, B_0) = 0.59$ while the MSE between the true and Bayesian maximum a posteriori was $mse(B_T, \tilde{B}) = 0.57$.

The maximum a posteriori variance for the source reference function was $\tilde{r}^2 = 0.30$, reduced from its prior mean of $(3/4)^2$ toward its true value of $(\frac{1}{4})^2$. The maximum a posteriori values of the observation error variances are

$$\tilde{\psi}^2 : \begin{pmatrix} 89.87 & 76.51 & 94.91 & 87.00 \\ 89.58 & 76.47 & 85.78 & 87.29 \\ 81.95 & 101.77 & 84.97 & 86.82 \\ 108.62 & 105.58 & 81.11 & 69.78 \end{pmatrix}.$$

The detrended time course for voxel 16 along with the Bayesian and prior square reference functions are given in Figure 3. Here detrended means both the linear trend was subtracted off and division of the appropriate mixing coefficient was performed. Note the similarity between the detrended observed time course and the Bayesian one. To quantify this similarity, the MSE's were computed to be $mse(s_0, s_T) = 0.25$ and $mse(\tilde{s}, s_T) = 0.15$. Note the smaller MSE for the Bayesian maximum a posteriori reference function.

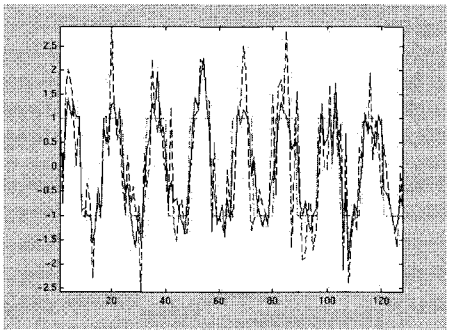

FIGURE 3. Detrended time course for voxel 16 $--$ with Prior Square, $-\cdot$ and Bayesian, — Reference Functions.

The Bayesian posterior mixing coefficients were found to be

$$\tilde{\Lambda} : \begin{pmatrix} 12.49 & 2.29 & 1.11 & 0.96 \\ 0.65 & 12.03 & -4.94 & 1.47 \\ -0.30 & -2.99 & 13.13 & 1.87 \\ 0.66 & 0.20 & 2.39 & 12.20 \end{pmatrix}.$$

The MSE between the true mixing coeficients and those from standard regression was $mse(\Lambda_T, \Lambda_0) = 7.70$ while the MSE between the true and Bayesian maximum a posteriori estimate was $mse(\Lambda_T, \tilde{\Lambda}) = 0.25$. The Bayesian values are much closer to the true values than those from multiple regression.

The correlation was computed between the square reference function and each of the respective detrended observed time courses.

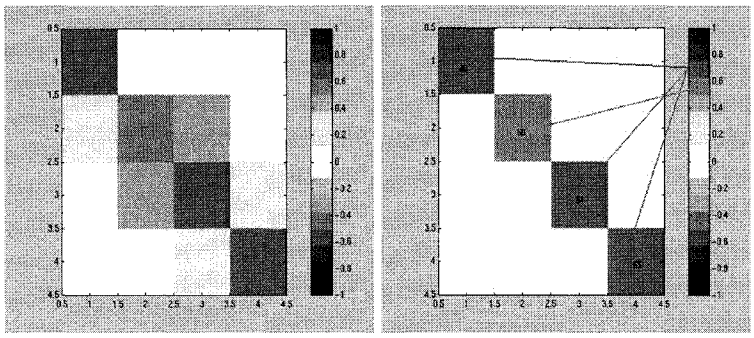

FIGURE 4. Activations for Prior Reference Function. Right thresholded at 0.5.

$$cc(x, s_0) = \begin{pmatrix} 0.66 & 0.01 & -0.03 & 0.06 \\ 0.23 & 0.60 & -0.26 & 0.06 \\ 0.05 & -0.36 & 0.64 & 0.24 \\ 0.10 & 0.08 & 0.14 & 0.65 \end{pmatrix}$$

As by design, large positive correlations are along the diagonal from upper left to lower right. A threshold was set at 0.5 and illustrated in Figure 4.

Upon computing the correlations between the Bayesian reference function and each of the respective detrended observed time courses.

$$cc(x, \tilde{s}) = \begin{pmatrix} 0.79 & 0.08 & -0.03 & 0.06 \\ 0.25 & 0.81 & -0.28 & 0.01 \\ 0.12 & -0.47 & 0.82 & 0.25 \\ 0.10 & 0.16 & 0.20 & 0.82 \end{pmatrix}$$

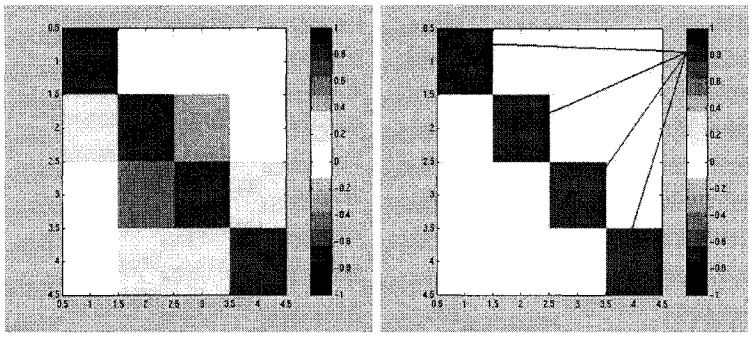

FIGURE 5. Activations for Bayesian Reference Function. Right thresholded at 0.5.

Applying the same threshold, all diagonal activations are present and more pronounced than those by correlation with the the standard regression method. The correlation along the diagonal of the image increased by an average of 0.17.

These functional activations are to be superimposed onto the previously shown anatomical image. In this example, the Bayesian statistical source separation model outperformed the common method of multiple regression.

It can be seen that the Bayesian method of determining the reference function for computation of voxel activation performed well especially with only sixteen voxels. In the next example with real fMRI data there are 98304 voxels.

REAL FMRI EXAMPLE

The current fMRI data provides the motivation for using Bayesian source separation to determine the true underling unobserved source reference function. This reference function is the underlying response due to the experimental stimulus. The data was collected from an experiment in which a subject was given eight rounds of tasks A and B. The timing was exactly the same as in the simulated example. The tasks were each 32 seconds in length and eight rounds were given. Task A was a complex task in which the subject read text from a screen, determined a number, entered the number using button response unit all in 22 seconds then received feedback displayed on a screen for 10 seconds. Task B was blank screen for 32 seconds.

For the functional data, 24 axial slices of size 64×64 were taken. Observations were taken every four seconds so that there are 128 in each voxel. All hyperparameters were assessed according to the empirical Bayes regression approach in Appendix A except for the prior mean and variance for the reference function. For the prior mean, a square wave was assessed with unit amplitude and frequency $1/64$ Hz which mimics the experiment. Due to space limitations, the prior and posterior parameter values for the 98304 voxels have been omitted.

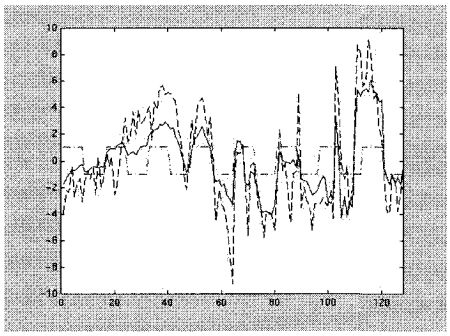

FIGURE 6. One voxels detrended time course $--$, the prior square $-\cdot$, and Bayesian $—$, reference functions.

One voxels detrended time course along with the prior square and Bayesian reference functions are given in Figure 6. Here detrended means both the linear trend was subtracted off and division of the appropriate mixing coefficient was performed. This voxel is located in the center of the activations shown in Figures 7 and 8. Note the similarity between the true time course and the Bayesian one. The correlation between this

detrended time course and the prior square wave was 0.28 while it was 0.86 with the Bayesian reference function.

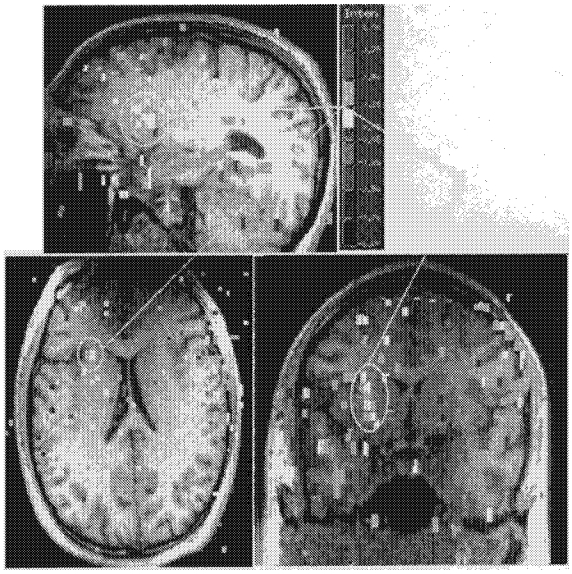

FIGURE 7. Activations for prior reference function thresholded at 0.22.

The correlation was computed between the square reference function and each of the respective detrended observed time courses. It is evident that the activation in Figure 7 is not very large and is buried in the noise. The threshold is set at 0.22 and if raised, the activation begins to disappear while noise remains.

The correlation was computed between the Bayesian underlying reference function and each of the respective detrended observed time courses. It is evident that the activation in Figure 8 is much larger and is no longer buried in the noise. The activations stand out above the noise. The threshold is set at 0.79 and if raised, the activation slowly begins to disappear.

The activations that were computed using the underlying reference function from Bayesian source separation were much larger and more distinct than those using a square reference function.

CONCLUSION

In computing the activations in fMRI, the choice of the reference function is subjective. It has been shown that the reference function need not be assigned but may be determined statistically using Bayesian methods. Further, this Bayesian reference function was more successful at identifing activations.

In certain fMRI applications, the experimental tasks are not simple, but consist of complex tasks that do not fit into the simple periodic framework. Determining a Bayesian reference function in this manner should prove to be useful for researchers. Work in

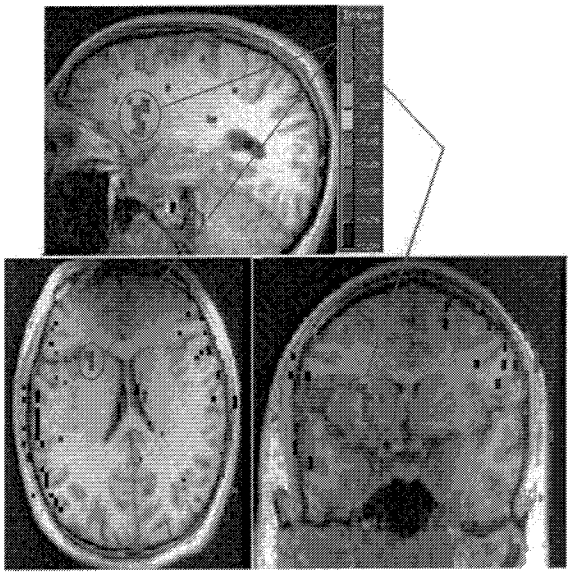

FIGURE 8. Activations for Bayesian reference function thresholded at 0.79.

progress includes taking other sources into account such as an EKG or respiration which will be columns of U. This statistical analysis will be able to more correctly identify significant voxels associated with the experimental task.

A. HYPERPARAMETER ASSESSMENT

The reference function corresponding to the experimental task is chosen to mimic the experiment with peaks during the experimental task and valleys during the control task. Typically a sine, square, or triangle wave function with unit ampitude and the same period as the experiment. Other source reference functions are assessed from a substantive field expert or possibly from an EKG or respiration monitor.

Reparameterizing the prior source matrix in terms of columns instead of rows as $S_0 = (s_{01}^*, \ldots, s_{0n}^*)$, each of these column vectors is the time course associated with a source reference function. Without loss of generality, denote the first as the one associated with the experimental tasks.

It is specified that a priori the covariance matrices are diagonal but free to be a posteriori non-diagonal. The hyperparameters for the covariance matrices are $A = a_0 I_p$, $v = v_0 I_m$, and $Q = q_0 I_p$. The hyperparameters v_0 and η are assessed as follows. From the prior distribution for R, the mean and variance of any diagonal element k which are a priori equal is

$$E(r_{kk}) = \frac{v_0}{\eta - 2m - 2}, \quad var(r_{kk}) = \frac{2v_0^2}{(\eta-2m-2)^2(\eta-2m-4)}. \tag{A.1}$$

The above is a system of two equations with two unknowns. Solving for η and v_0 yields

$$\eta = \frac{[E(r_{kk})]^2}{2var(r_{kk})} + 6, \quad v_0 = E(r_{kk})(\eta - 4). \tag{A.2}$$

Similarly, the hyperparameters ν and b_0 for the error of observation are

$$\nu = \frac{[E(\psi^2)]^2}{2var(\psi^2)} + 6, \quad q_0 = E(\psi^2)(\eta - 4). \tag{A.3}$$

The problem of assessing hyperparameters for these distributions is now transformed to assessing prior means and variances. This may be done subjectively by an expert, from a previous data set, or in an empirical Bayes approach from the current data set.

The hyperparameter Λ_0 can be assessed from an expert or in the following way from previous or current data. Perform a regression of the observed data on the assessed prior means of the source reference functions. That is, find the λ_{0jk}'s as regression coefficients in

$$x_{ij} = \beta'_{jT} u_i + \lambda'_j s_i + \varepsilon_{ij}. \tag{A.4}$$

For each voxel j, there will be an estimate of the error variance

$$\hat{\psi}_j^2 = \frac{1}{n} \sum_{i=1}^{n} \left(x_{ij} - \beta'_j u_i - \hat{\lambda}'_j s_{0i} \right)^2. \tag{A.5}$$

and an estimate of each of the variances of the regression coefficients $\hat{r}_{jkk}$. The prior means and variances $E(\psi^2)$, $var(\psi^2)$, $E(r_{kk})$, and $var(r_{kk})$ may be estimated as the sample means and variances from these regressions. In addition, the hyperparameter h_0 may be estimated as $a_0 = \hat{E}(r_{kk})/\hat{E}(\psi^2)$, the quotient of these estimated mean variances.

REFERENCES

1. P. Bandetti, A. Jesmanowicz, E. Wong, and J. Hyde, "Processing strategies for time-course data sets in functional mri of the human brain," *MRM*, **30**, 1993.
2. R. W. Cox, A. Jesmanowicz, and J. Hyde, "Real-time functional magnetic resonance imaging," *MRM*, **33**, 1995.
3. K. Knuth, "Bayesian source separation and localization," in *SPIE'98 Proceedings: Bayesian Inference for Inverse Problems, San Diego, CA*, A. Mohammad-Djafari, ed., pp. 147–158, July 1998.
4. K. Knuth, "A Bayesian approach to source separation," in *Proceedings of the First International Workshop on Independent Component Analysis and Signal Separation: ICA'99, Aussios, France*, C. J. J.-F. Cardoso and P. Loubaton, eds., pp. 283–288, 1999.
5. D. B. Rowe, "A Bayesian approach to blind source separation," Journal of Interdisciplinary Mathematics. To appear.
6. D. V. Lindley and A. F. M. Smith, "Bayes estimates for the linear model," *Journal of the Royal Statistical Society B*, **34**, (1), 1972.
7. A. O'Hagen, *Kendalls' Advanced Theory of Statistics, Volume 2B Bayesian Inference*, John Wiley and Sons Inc., 1994.

8. D. B. Rowe and S. J. Press, "Gibbs sampling and hill climbing in Bayesian factor analysis," Tech. Rep. No. 255, Department of Statistics, University of California, Riverside, May 1998.
9. A. Mohammad-Djafari, "A Bayesian estimation method for detection, localisation and estimation of superposed sources in remote sensing," in *SPIE 97 annual meeting, (San Diego, CA, USA, July 27-Aug. 1, 1997)*, 1997.
10. A. Mohammad-Djafari, "A Bayesian approach to source separation," in *Proceedings of The Nineteenth International Conference on Maximum Entropy and Bayesian Methods, (August 2-6, 1999, Boise, ID.)*, 1999.
11. R. W. Cox, "AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages," *Computers and Biomedical Research*, **29**, 1996.