Bayesian Source Separation for Reference Function Determination in fMRI

Daniel B. Rowe*

In analyzing fMRI results, identification of significant activation in voxels is a crucial task. A standard method selects a "known" reference function and performs a 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. But the most important question is: How does one choose the reference function? Here, a Bayesian source separation approach to determining the underlying reference function is described and applied to real fMRI data. This underlying reference function is the unobserved response due to the presentation of the experimental stimulus. Magn Reson Med 46: 374–378, 2001. © 2001 Wiley-Liss, Inc.

Key words: fMRI; Bayesian; reference; function

Typically in an fMRI, a sequence of two stimuli or tasks, A and B, are given to participants. Imaging takes place while the participant is responding either passively or actively to these tasks. A model is used to describe the observed signal in each voxel 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. Levels of activation are assigned coloration and voxels colored accordingly.

In computing a voxel's activation, a standard method (1,2) is to correlate the detrended observed time-course in each voxel and an assumed to be known reference function. But the most important question is: How does one choose the reference function? This article develops a coherent Bayesian statistical approach to determine the underlying response or reference function rooted in Bayesian source separation. The reference function is viewed as the underlying response due to the presentation of the experimental tasks. The reference function need not fit into the standard on/off or rise/fall format and it may change (possibly nonlinearly) over the course of the experiment. In this approach, all the voxels contribute to "telling us" the underlying response due to the presented experimental stimulus.

In the source separation model (3), observations consist of mixtures of true unobservable signals. 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. 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 also allows for several different and possibly correlated component time sequences. There are always incidental cognitive processes and blood flow that may be considered components. These time sequences could correspond to cardiac and respiration activity. Modeling them instead of grouping them into the error term could prove useful.

THEORY

Model and Likelihood

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

$$x_{ij} = g_j + h_j i + l_{j1}s_{i1} + \cdots + l_{jm}s_{im} + e_{ij}. \quad [1.1]$$

The observed signal in voxel j at time i contains a linear part with a voxel specific 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 $l_{j1},\ldots,l_{jm}$. This model can be written in terms of vectors as:

$$x_{ij} = \mathbf{b}_j'\mathbf{u}_i + \mathbf{l}_j'\mathbf{s}_i + e_{ij}. \quad [1.2]$$

where $\mathbf{u}_i = (1,i)'$, $\mathbf{b}_j = (g_j,h_j)'$, $\mathbf{l}_j = (l_{j1},\ldots,l_{jm})'$ and $\mathbf{s}_i = (s_{i1},\ldots,s_{im})'$. 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.

Considering all p voxels at time i, the model can be written as:

$$\mathbf{x}_i = \mathbf{B}\mathbf{u}_i + \mathbf{L}\mathbf{s}_i + \mathbf{e}_i \quad [1.3]$$

Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, California.

*Correspondence to: Daniel B. Rowe, Division of Humanities and Social Sciences, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125. E-mail: drowe@hss.caltech.edu

Received 6 June 2000; revised 18 January 2001; accepted 5 March 2001.

where $\mathbf{B} = (\mathbf{b}_1,\ldots,\mathbf{b}_p)'$, is a $p \times 2$ matrix of slopes and intercepts, and $\mathbf{L} = (\mathbf{l}_1,\ldots,\mathbf{l}_p)'$ is a $p \times m$ dimensional matrix of mixing coefficients.

The model which considers all voxels at all time points written in terms of matrices is:

$$\mathbf{X} = \mathbf{U}\mathbf{B}' + \mathbf{S}\mathbf{L}' + \mathbf{E} \quad [1.4]$$

where $\mathbf{X}' = (\mathbf{x}_1,\ldots,\mathbf{x}_n)$ is a $p \times n$ matrix of observations, $\mathbf{U} = (\mathbf{o}_n,\mathbf{c}_n)$ is a $n \times 2$ matrix, $\mathbf{o}_n$ is a $n \times 1$ vector of ones, $\mathbf{c}_n = (1,\ldots,n)'$, $\mathbf{S}' = (\mathbf{s}_1,\ldots,\mathbf{s}_n)$ is a $m \times n$ matrix of unobservable sources, and $E' = (\mathbf{e}_1,\ldots,\mathbf{e}_n)$ is a $p \times n$ matrix of errors while $\mathbf{B}$ and $\mathbf{L}$ are as before.

Motivated by the central limit theorem, the errors of observation are taken to be normally distributed, with mean zero and covariance $\mathbf{P}$ yielding the normal likelihood:

$$p(\mathbf{X}|\mathbf{B},\mathbf{S},\mathbf{L},\mathbf{P}) \propto |\mathbf{P}|^{-(n/2)}\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{X}-\mathbf{U}\mathbf{B}'-\mathbf{S}\mathbf{L}')\mathbf{P}^{-1} \times (\mathbf{X}-\mathbf{U}\mathbf{B}'-\mathbf{S}\mathbf{L}')'\right\}. \quad [1.5]$$

Priors and Posterior

Recall that the source components, the $\mathbf{s}_i$'s, are unobserved. As previously noted, in fMRI 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 in each voxel to fit the model:

$$x_{ij} = g_j + h_j i + \mathbf{l}_j\mathbf{s}_i + \mathbf{e}_{ij} \quad [2.1]$$

and obtain the regression coefficient estimates $(\hat{g}_j, \hat{h}_j, \hat{l}_j)$.

Significant activation is determined by correlation and voxels are assigned coloration according to their activation level.

The above method of subjectively assigning 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 to 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, then correlation is performed. In addition, prior distributions are assessed for the covariance matrix for the sources, the slopes and intercepts, the mixing coefficients, and the covariance matrix for the observation error.

When quantifying available prior information (4) 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 (the multivariate version of the Inverted Gamma). A vague or noninformative distribution is taken for the regression coefficients.

The prior distributions for the parameters of interest are:

$$p(\mathbf{S}|\mathbf{R}) \propto |\mathbf{R}|^{-(n/2)}\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{S}-\mathbf{S}_0)\mathbf{R}^{-1}(\mathbf{S}-\mathbf{S}_0)'\right\} \quad [2.2]$$

$$p(\mathbf{R}) \propto |\mathbf{R}|^{-(d/2)}\exp\left\{-\frac{1}{2}\mathrm{tr}\mathbf{R}^{-1}\mathbf{V}\right\} \quad [2.3]$$

$$p(\mathbf{L}|\mathbf{P}) \propto |\mathbf{A}|^{-(p/2)}|\mathbf{P}|^{-(m/2)} \times \exp\left\{-\frac{1}{2}\mathrm{tr}A^{-1}(\mathbf{L}-\mathbf{L}_0)'\mathbf{P}^{-1}(\mathbf{L}-\mathbf{L}_0)\right\} \quad [2.4]$$

$$p(\mathbf{P}) \propto |\mathbf{P}|^{-(f/2)}\exp\left\{-\frac{1}{2}\mathrm{tr}\mathbf{P}^{-1}\mathbf{Q}\right\}, \quad [2.5]$$

$$p(\mathbf{B}) \propto \text{constant} \quad [2.6]$$

where the prior mean for the source reference functions is $\mathbf{S}'_0 = (\mathbf{s}_{01},\ldots,\mathbf{s}_{0n})$, $\mathbf{R}$ is the $m \times m$ source covariance matrix; $\mathbf{V}$ an $m \times m$ matrix along with d define the prior for the source covariance matrix; $\mathbf{A}$ is a hyperparameter defining the prior for the mixing matrix; while $\mathbf{P}$ is the $p \times p$ covariance matrix for the errors whose prior distribution is defined by $\mathbf{Q}$ a $p \times p$ matrix along with f. The hyperparameters $\mathbf{S}_0$, d, $\mathbf{V}$, $\mathbf{A}$, f, and $\mathbf{Q}$ are to be assessed.

The data from all p time courses of length n will be combined with the prior distributions to produce a joint posterior distribution. Maximum a posteriori estimates can be obtained from the joint posterior via the iterated conditional modes (ICM) algorithm (5–7) which iterates through the modes of the posterior conditional distributions.

The ICM algorithm consists of starting with initial $\mathbf{S}$ and $\mathbf{B}$ values, say $\tilde{\mathbf{S}}_{(0)}$ and $\tilde{\mathbf{B}}_{(0)}$, then iterating through:

$$\tilde{\mathbf{L}}_{(k+1)} = [\mathbf{L}_0\mathbf{A}^{-1} - (\mathbf{X}-\mathbf{U}\tilde{\mathbf{B}}_{(k)}')'\tilde{\mathbf{S}}_{(k)}](\mathbf{A}^{-1} + \tilde{\mathbf{S}}_{(k)}'\tilde{\mathbf{S}}_{(k)})^{-1} \quad [2.7]$$

$$\tilde{\mathbf{P}}_{(k+1)} = \frac{1}{n+m+f}\{(\mathbf{X}-\mathbf{U}\tilde{\mathbf{B}}_{(k)}' - \tilde{\mathbf{S}}_{(k)}\tilde{\mathbf{L}}_{(k+1)}')'(\mathbf{X}-\mathbf{U}\tilde{\mathbf{B}}_{(k)}' - \tilde{\mathbf{S}}_{(k)}\tilde{\mathbf{L}}_{(k+1)}') + (\tilde{\mathbf{L}}_{(k+1)}-\mathbf{L}_0)\mathbf{A}^{-1}(\tilde{\mathbf{L}}_{(k+1)}-\mathbf{L}_0)' + \mathbf{Q}\} \quad [2.8]$$

$$\tilde{\mathbf{R}}_{(k+1)} = \frac{(\tilde{\mathbf{S}}_{(k)}-\mathbf{S}_0)'(\tilde{\mathbf{S}}_{(k)}-\mathbf{S}_0) + V}{n+d} \quad [2.9]$$

$$\tilde{\mathbf{S}}_{(k+1)} = [\mathbf{S}_0\tilde{\mathbf{R}}_{(k+1)}^{-1} + (\mathbf{X}-\mathbf{U}\tilde{\mathbf{B}}_{(k)}')\tilde{\mathbf{P}}_{(k+1)}^{-1}\tilde{\mathbf{L}}_{(k+1)}] \times (\tilde{\mathbf{R}}_{(k+1)}^{-1} + \tilde{\mathbf{L}}_{(k+1)}'\tilde{\mathbf{P}}_{(k+1)}^{-1}\tilde{\mathbf{L}}_{(k+1)})^{-1} \quad [2.10]$$

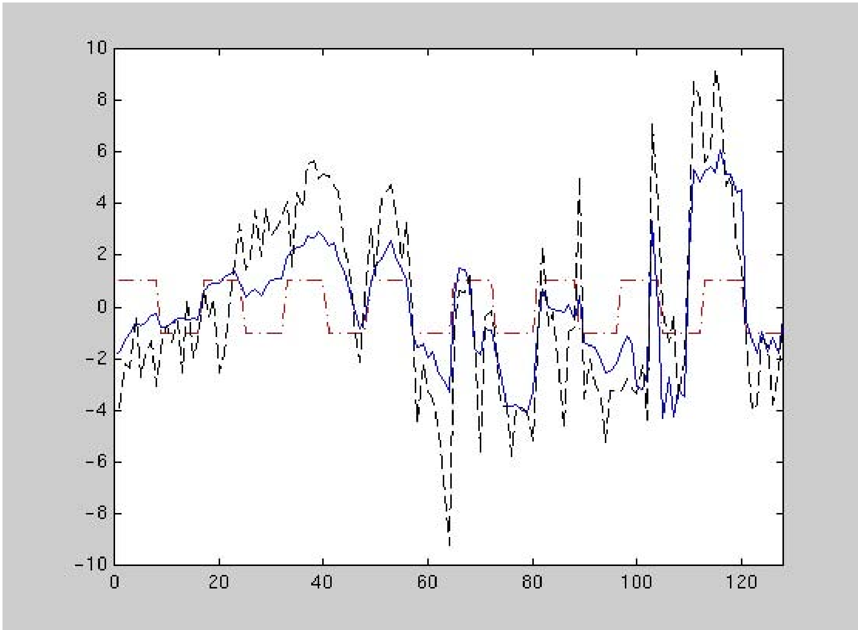

FIG. 1. One voxel's detrended time course —, the prior square –·, and Bayesian —, reference functions. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

$$\tilde{\mathbf{B}}_{(k+1)} = (X - \tilde{\mathbf{S}}_{(k+1)}\tilde{\mathbf{L}}_{(k+1)}{}')'\mathbf{U}(\mathbf{U}'\mathbf{U})^{-1} \qquad [2.11]$$

until convergence is reached.

With the typical number of voxels, there are an enormous number of distinct covariance elements. Implementation is impractical because **P** is inverted in the ICM algorithm. It is currently assumed for computational purposes only and not a model assumption that the voxels are spatially uncorrelated. This assumption could be relaxed by considering local correlation.

METHODS

The current fMRI study, which obtained the patients informed consent and complied with the institution's ethics committee, provides the motivation for using Bayesian source separation to determine the true underling unobserved source reference function, which is the underlying response due to the experimental stimulus. The data were collected from an experiment in which a participant was given eight rounds of tasks A and B. The tasks were each 32 sec in length. Task A was complex, which consisted of several subtasks and is not well described by a square reference function. Task B consisted of a blank screen. Had task A been a simple repetitive task, then the known square reference function would be reasonable.

Experimental task A was an implementation of a common experimental economic method for determining participants' valuation of objects in the setting of an auction (8). The participant was given an item and told that the item can be kept or sold. If the item is kept, the participant retains ownership at the end of the experiment and is paid the items stated value.

If the item is sold, then the participant receives the selling price. The participant's task is to name the lowest price at which the participant is willing to sell the item. That price will be compared with an offer price randomly generated independently of the participant's price. If the participant's is lower than the random offer price, the participant sells the item at the offer price. Otherwise the participant keeps the item. The best price for the participant to ask for is the participant's true valuation of the item. Stating any other price leads to the risk of either selling the item below its value or of rejecting a bid above the value. In the experiment the participant has to learn the strategy in order to maximize the earnings.

The method was implemented using coupons that could be redeemed for cash at the end of the experiment. The participant was given the right to a coupon with a stated value. The participant was asked to enter their selling price for the coupon. The participant's selling price is compared to a predetermined randomly generated offer. If the offer is above the participant's selling price, then the participant sells the coupon and earns the offer price; otherwise, the participant keeps the coupon and is paid its value. The participant is shown bids, whether or not a sale was made, and the amount the participant earned on that repetition. The total earnings are shown and updated on each repetition. While in the scanner the participant communicates bids using a specially designed nonmagnetic fiber optic Button Response Unit and receives instructions from images projected into the scanner. Task A consisted of the participant viewing the screen, determining his/her valuation, then entering the valuation and receiving feedback as to her/his performance through earnings.

For the functional data, 24 axial slices of size 64 × 64 were taken. Each voxel had dimensions of 3 × 3 × 5 mm. Scanning was performed using a 1.5 T Siemens Magneton with TE = 40 ms. Observations were taken

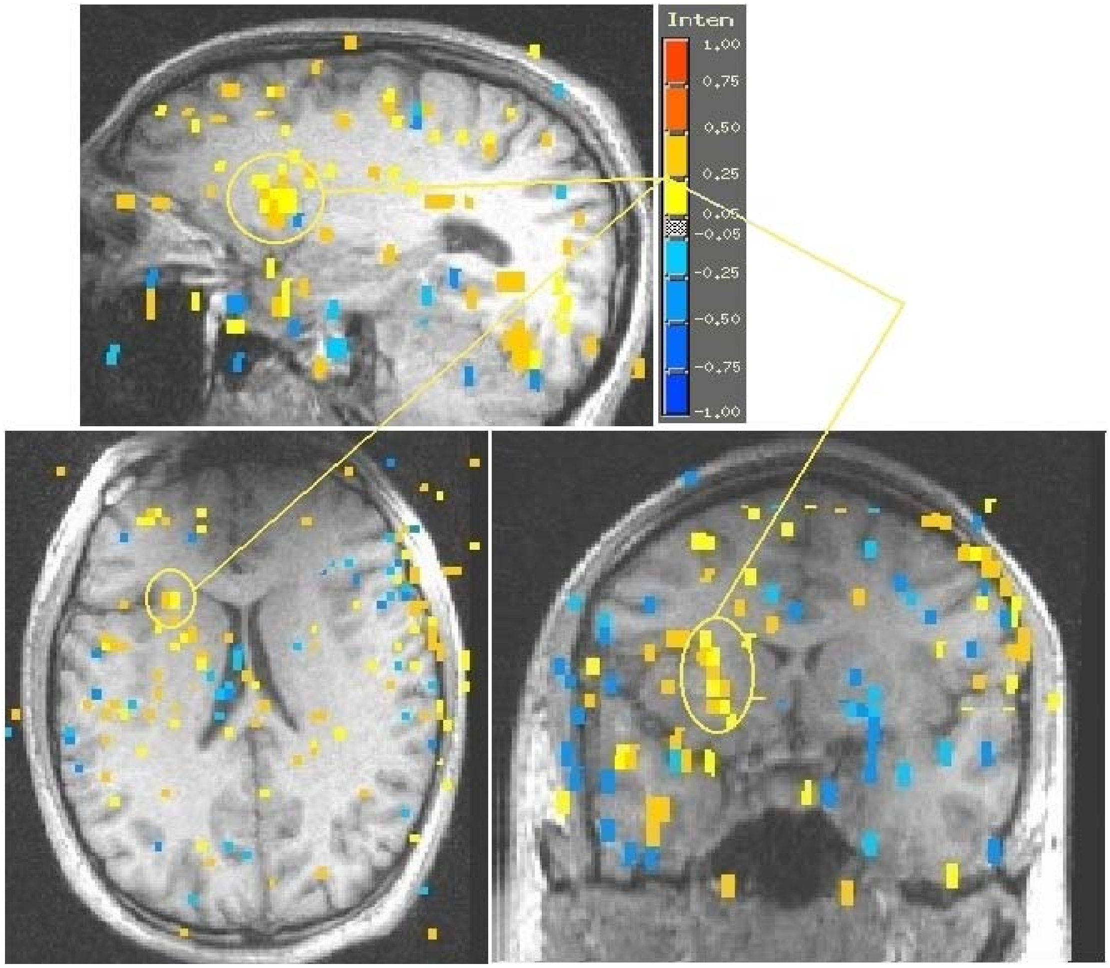

FIG. 2. Activations for (**a**) prior thresholded at 0.22 and (**b**) Bayesian at 0.79 reference functions. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

every 4 sec so that there are 128 in each voxel. All hyperparameters were assessed according to an empirical Bayes regression approach which fit a multiple regression model to the data 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.

RESULTS

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

One voxel's detrended time course along with the prior square and Bayesian reference functions are given in Fig. 1. This voxel is located in the center of a group of activations located in the anterior putamen, as shown in Fig. 2. Here detrended means both the linear trend was subtracted off and division of the appropriate mixing coefficient was performed.

Note the similarity between the true time course and the Bayesian reference function. 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. The correlation was computed between the reference function and each of the respective detrended observed time courses. It is evident that the circled activation in Fig. 2a 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 circled activation in Fig. 2b appears larger and no longer buried in the noise. The threshold is set at 0.79 and if raised the activation slowly begins to disappear.

DISCUSSION

In computing the activations in fMRI, the choice of the reference function is subjective. It has been shown that the

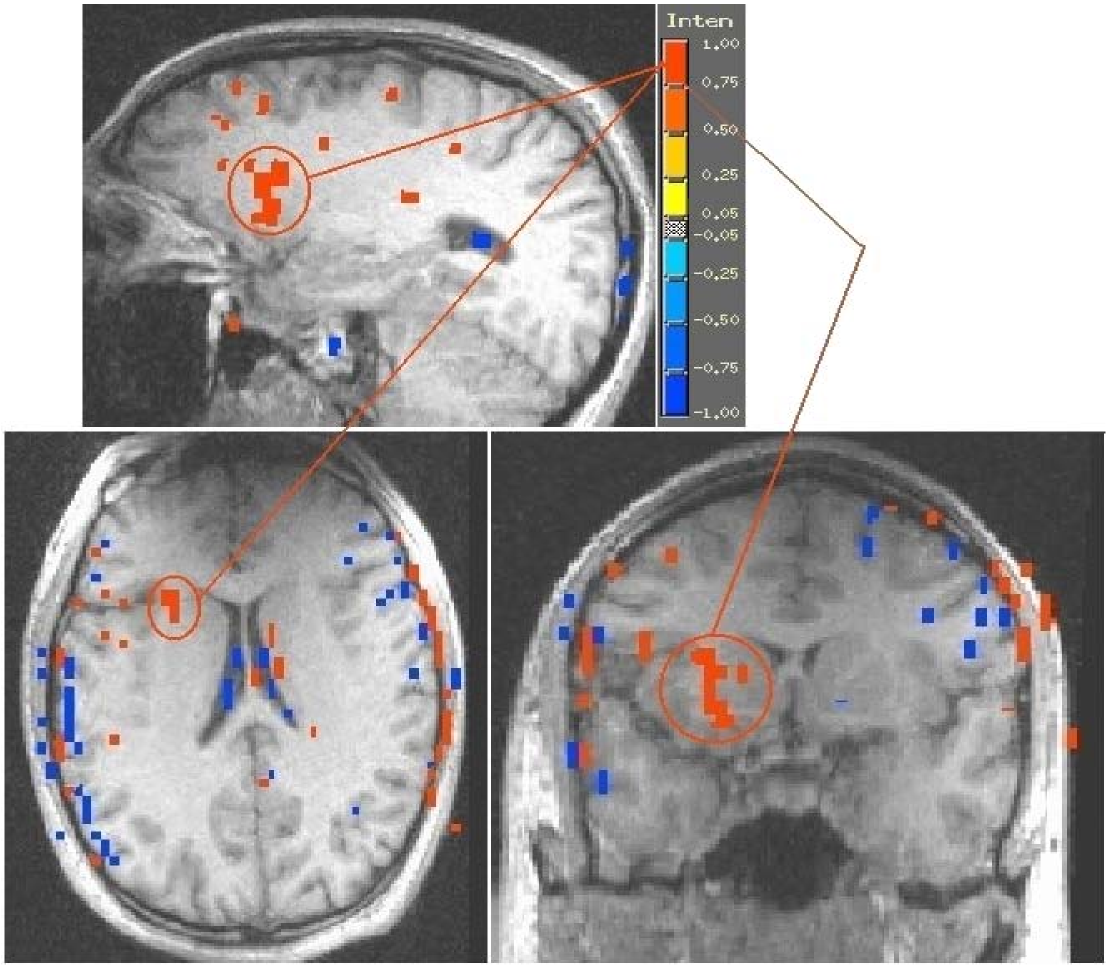

FIG. 2. Continued.

reference function need not be assigned but may be determined statistically using Bayesian methods. Further, this Bayesian reference function produced more distinct 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.

REFERENCES

1. Bandettini P, Jesmanowicz A, Wong E, Hyde J. Processing strategies for time-course data sets in functional MRI of the human brain. Magn Reson Med 1993;30:161–173.
2. Cox RW, Jesmanowicz A, Hyde J. Real-time functional magnetic resonance imaging. Magn Reson Med 1995;33:230–236.
3. Knuth K. Bayesian source separation and localization. In: Mohammad-Djafari A, editor. In: Proceedings of SPIE'98: Bayesian Inference for Inverse Problems, San Diego, 1998.
4. Press SJ. Applied multivariate analysis: using Bayesian and frequentist methods of inference. Malabar, FL: Robert E. Krieger Publishing; 1982.
5. Lindley DV, Smith AF. Bayes estimates for the linear model. J R Stat Soc B 1972;34:1–19.
6. O'Hagen A. Kendalls' advanced theory of statistics, vol. 2B, Bayesian inference. New York: John Wiley & Sons; 1994.
7. Rowe DB, Press SJ. Gibbs sampling and hill climbing in Bayesian factor analysis. University of California, Riverside, Department of Statistics, Technical report 255.
8. Becker GM, De Groot MH, Marshak M. Measuring utility by a single response sequential method. Behav Sci 1964;9:226–232.