Real-Time Functional Magnetic Resonance Imaging

Robert W. Cox, Andrzej Jesmanowicz, James S. Hyde

A recursive algorithm suitable for functional magnetic resonance imaging (FMRI) calculations is presented. The correlation coefficient of a time course of images with a reference time series, with the mean and any linear trend projected out, may be computed with 22 operations per voxel, per image; the storage overhead is four numbers per voxel. A statistical model for the FMRI signal is presented, and thresholds for the correlation coefficient are derived from it. Selected images from the first real-time functional neuroimaging experiment (at 3 Tesla) are presented. Using a 50-MHz workstation equipped with a 14-bit analog-to-digital converter, each echo planar image was acquired, reconstructed, correlated, thresholded, and displayed in pseudocolor (highlighting active regions in the brain) within 500 ms of the RF pulse.

Key words: functional MRI; recursive image processing.

INTRODUCTION

In the last few years, MRI has demonstrated the ability to detect changes in cerebral blood volume (1), blood flow (2), and blood oxygenation (3, 4) that occur locally in association with increased neuronal activity. The most widely used MRI method for the noninvasive mapping of human brain activity is based on blood oxygenation level dependent (BOLD) contrast (5). Functional MRI (FMRI) has become a highly used technique for noninvasive mapping and analysis of cortical activity in humans (6).

The percent signal change in activated brain regions is small, even at high fields (2, 3). Reliable mapping of brain activity thus requires gathering a large number of images in both the active and inactive states, so that the noise may be reduced by averaging (7). Here, "noise" includes not only the usual MRI sensor noise but also physiological fluctuations that affect the signal (8–10).

A typical FMRI scanning session generates a large amount of data—hundreds to tens of thousands of images. At present, image combination to extract useful functional information is done after the data acquisition is complete. Because scanner time is expensive, and it is unreasonable to ask a patient or experimental volunteer to wait while data analysis software is run, it is quite common to postpone functional data "postprocessing" until the scanning session is completed.

MRM 33:230–236 (1995)

From the Biophysics Research Institute, Medical College of Wisconsin, Milwaukee, Wisconsin.

Address correspondence to: Robert W. Cox, Ph.D., Biophysics Research Institute, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226–0509.

Received July 5, 1994; revised October 12, 1994; accepted October 12, 1994.

This work was supported in part by grant CA41464 from the National Institutes of Health.

0740-3194/95 $3.00

There are three principal reasons why a capability for near-real-time viewing of the FMRI activation calculations is desirable:

1. Data quality may be monitored as an experiment progresses; the investigator will know if functional activation was measured during an experiment, instead of waiting to postprocess the data. In particular, the effects of stimulus-correlated motion (11), which are spread throughout the field-of-view (FOV) at all areas with strong MR signal intensity gradients, may be observed immediately, enabling the investigator to reacquire the contaminated data set immediately.
2. Real-time FMRI will make it possible to develop new task and stimulus protocols much more quickly than can be done with "day-after" data analyses.
3. Interactive experimental paradigms may be created, making FMRI a more flexible tool for neurological investigations. For such applications to be possible, the ability to see the progress of a scanning session as it occurs will be essential.

Real-time FMRI processing and display will never be a complete substitute for postprocessing, where multiple statistical methods may be applied to a data set for detection of activation and suppression of artifacts. In general, FMRI data have a rich structure and are susceptible to many artifacts. The results of each experiment should be examined from a number of points of view by a trained investigator.

The algorithms designed for processing an entire set of images at once ("batch processing") are not usually suitable for "real-time" processing (i.e., quickly analyzing incomplete and growing data sets). A powerful method for combining image sequences into an activation map is the correlation technique of Bandettini *et al.* (7), where the time series in each voxel is cross-correlated with a "reference vector." The straightforward implementation of this method requires that all the images be gathered before processing starts.

We present an algorithm that recursively computes the correlation coefficient of an image sequence with the reference vector, and simultaneously projects out (in a least squares sense) any undesired time series (e.g., a linear trend). The additional computational load for the FMRI calculations is minimal if real-time image reconstruction is available. We also present an analysis of the statistics of the threshold test, derived from a signal model. Results from real-time echo-planar FMRI data acquisition and processing on a 3 T scanner are shown.

PARTIAL CORRELATION COEFFICIENT AND FMRI

The correlation method in FMRI is best expressed in a vector space formalism. If $\mathbf{x}$ is the vector formed from the time series of MR image intensities in a single voxel, and

$\mathbf{r}$ is a "reference" vector, then the correlation coefficient between $\mathbf{x}$ and $\mathbf{r}$ is

$$\rho = \frac{\mathbf{r}^T\mathbf{x}}{|\mathbf{r}| \cdot |\mathbf{x}|} \quad \alpha = \rho \cdot \frac{|\mathbf{x}|}{|\mathbf{r}|} = \frac{\mathbf{r}^T\mathbf{x}}{|\mathbf{r}|^2}.$$

(Mathematical notations and definitions are gathered together in the Appendix.) Here, α is the coefficient that minimizes $|\mathbf{x} - \alpha\mathbf{r}|$; it provides the best fit (in the least squares sense) of $\mathbf{x}$ to $\mathbf{r}$. By the Cauchy-Schwarz inequality, $-1 \le \rho \le 1$.

In the application to FMRI, the reference $\mathbf{r}$ is derived from the timing of the mental tasks being performed by the subject. The simplest case is that of an *"on:off:on:off"* task paradigm; a square wave (equal to 1 during *"on"* intervals and equal to 0 during *"off"* intervals) is an obvious choice for $\mathbf{r}$. Voxels whose signal levels change synchronously with the task will have a large ρ; those whose signal levels simply fluctuate randomly will have a small ρ. We declare as activated those voxels with $|\rho| \ge \rho_{\text{thr}}$. The choice of the threshold ρ_{thr} is discussed in the Statistics of ρ section. We use the amplitude α as a measure of the strength of the signal changes in each voxel that is above the threshold.

More complex methods for choosing $\mathbf{r}$ have been discussed at length (7). One crucial point is that it is necessary to remove the mean signal level from $\mathbf{r}$ and $\mathbf{x}$ before computing ρ, because the signal changes in BOLD contrast are relatively small. It may also be desirable to remove any linear trend in the data, which could be due to instrument drift or slow subject motion. If we denote by $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_L\}$ the set of vectors that we wish to remove from the data, then the resulting signal model is

$$\mathbf{x} = \alpha\mathbf{r} + \sum_{k=1}^{L} \gamma_k \mathbf{s}_k; \qquad [1]$$

that is, that the measured vector $\mathbf{x}$ in any voxel is a linear combination of a reference vector $\mathbf{r}$ and a finite set of "trends" $\{\mathbf{s}_k\}$ induced by experimental artifacts. To remove the mean and linear trend, for example, we take $L = 2$, $\mathbf{s}_1 = [1\ 1 \ldots 1]^T$ and $\mathbf{s}_2 = [1\ 2 \ldots N]^T$.

Denote by $\mathbf{P}$ the projection matrix that annihilates the vectors $\{\mathbf{s}_k\}$. Then better definitions of ρ and α are

$$\rho = \frac{(\mathbf{Pr})^T(\mathbf{Px})}{|\mathbf{Pr}| \cdot |\mathbf{Px}|} = \frac{\mathbf{r}^T\mathbf{Px}}{(\mathbf{r}^T\mathbf{Pr} \cdot \mathbf{x}^T\mathbf{Px})^{1/2}} \quad \alpha = \frac{\mathbf{r}^T\mathbf{Px}}{\mathbf{r}^T\mathbf{Pr}}. \qquad [2]$$

When signals are removed from $\mathbf{x}$ and $\mathbf{r}$ prior to computation of ρ, then ρ is called the "partial correlation coefficient" of $\mathbf{x}$ with $\mathbf{r}$ (12). The definitions in Eq. [2] will be derived from a statistical model in the Statistics of ρ section.

Equation [2] is not well suited for real-time FMRI. When one more image is acquired, to recompute ρ for each voxel requires updating the least squares removal of $\mathbf{S}$ from $\mathbf{x}$, which will almost certainly alter every element of $\mathbf{Px}$. The calculation of ρ will then require recomputation of the scalar products in the numerator and denominator of Eq. [2]. As the number of images grows (i.e., as the vectors increase in dimension), the amount of calculation will grow. In a real-time application, this is unacceptable, because at some point the computer will not be able to finish processing a new image before the next one is ready.

RECURSIVE CALCULATION OF ρ

The first step in the derivation of a recursive technique for updating ρ is to consider the $(L + 2) \times (L + 2)$ matrix

$$\tilde{\mathbf{S}}^T\tilde{\mathbf{S}} = \begin{bmatrix} \mathbf{S}^T \\ \mathbf{r}^T \\ \mathbf{x}^T \end{bmatrix} [\mathbf{S}|\mathbf{r}\ \mathbf{x}] = \left[\begin{array}{c|cc} \mathbf{S}^T\mathbf{S} & \mathbf{S}^T\mathbf{r} & \mathbf{S}^T\mathbf{x} \\ \hline \mathbf{r}^T\mathbf{S} & \mathbf{r}^T\mathbf{r} & \mathbf{r}^T\mathbf{x} \\ \mathbf{x}^T\mathbf{S} & \mathbf{x}^T\mathbf{r} & \mathbf{x}^T\mathbf{x} \end{array}\right].$$

This matrix contains the inner products of all vectors involved in the calculation of ρ. Calculation of the inverse of $\tilde{\mathbf{S}}^T\tilde{\mathbf{S}}$ using the bordering technique (13) shows that the lower right corner 2×2 submatrix of $[\tilde{\mathbf{S}}^T\tilde{\mathbf{S}}]^{-1}$ is

$$\begin{aligned} \mathbf{\Delta} &= \begin{bmatrix} \mathbf{r}^T\mathbf{Pr} & \mathbf{r}^T\mathbf{Px} \\ \mathbf{x}^T\mathbf{Pr} & \mathbf{x}^T\mathbf{Px} \end{bmatrix}^{-1} \\ &= \frac{1}{\mathbf{r}^T\mathbf{Pr} \cdot \mathbf{x}^T\mathbf{Px} - (\mathbf{r}^T\mathbf{Px})^2} \begin{bmatrix} \mathbf{x}^T\mathbf{Px} & -\mathbf{r}^T\mathbf{Px} \\ -\mathbf{x}^T\mathbf{Pr} & \mathbf{r}^T\mathbf{Pr} \end{bmatrix}. \end{aligned}$$

Thus, once this portion of $[\tilde{\mathbf{S}}^T\tilde{\mathbf{S}}]^{-1}$ is available, then ρ and α are computable via

$$\rho = -\frac{\Delta_{12}}{(\Delta_{11}\Delta_{22})^{1/2}} \quad \alpha = -\frac{\Delta_{12}}{\Delta_{22}}. \qquad [3]$$

In practice, we wish to update $\rho^{(m)}$ to $\rho^{(m+1)}$ when we add a new row into $\tilde{\mathbf{S}}^{(m)}$ to make it $\tilde{\mathbf{S}}^{(m+1)}$. Using the notation $\boldsymbol{\sigma}_{m+1}^T$ for this last row, we see that

$$[\tilde{\mathbf{S}}^{(m+1)}]^T[\tilde{\mathbf{S}}^{(m+1)}] = [\tilde{\mathbf{S}}^{(m)}]^T[\tilde{\mathbf{S}}^{(m)}] + \boldsymbol{\sigma}_{m+1}\boldsymbol{\sigma}_{m+1}^T;$$

that is, the change in $\tilde{\mathbf{S}}^T\tilde{\mathbf{S}}$ when a new image is available is a rank one update. If the inverse is available at time step m, then the Sherman-Morrison formula could be used to modify it without a full recalculation (14, p. 890). This, however, is not the most efficient way in which to proceed.

The Cholesky decomposition (15) of $\tilde{\mathbf{S}}^T\tilde{\mathbf{S}} = \mathbf{C}\mathbf{C}^T$ may also be used to evaluate ρ. This is more efficient because it is faster to compute the Cholesky factor $\mathbf{C}$ than the matrix inverse, and we only need the lower right corner elements of $[\tilde{\mathbf{S}}^T\tilde{\mathbf{S}}]^{-1}$ in Eq. [3]. Straightforward calculation shows that

$$\Delta_{11} = \frac{1}{c_{L+1,L+1}^2}\left[1 + \frac{c_{L+2,L+1}^2}{c_{L+2,L+2}^2}\right]$$

$$\Delta_{12} = -\frac{c_{L+2,L+1}}{c_{L+2,L+2}^2 c_{L+1,L+1}} \quad \Delta_{22} = \frac{1}{c_{L+2,L+2}^2}.$$

Using these formulas, we find expressions for ρ and α:

$$\rho = \frac{c_{L+2,L+1}}{(c_{L+2,L+2}^2 + c_{L+2,L+1}^2)^{1/2}} \quad \alpha = \frac{c_{L+2,L+1}}{c_{L+1,L+1}}$$

Given $\mathbf{C}^{(m)}$ and $\boldsymbol{\sigma}_{m+1}$, Carlson's algorithm is a method for updating the Cholesky factor into $\mathbf{C}^{(m+1)}$ (16). Furthermore, the first $L + 1$ elements of $\boldsymbol{\sigma}_{m+1}$ are just the detrending and reference vectors at time index $m + 1$, and so do not vary between voxels. This means that the first $L + 1$ rows of $\mathbf{C}^{(m)}$ are the same for all voxels, and so

only need to be stored and computed once. The last row of $\mathbf{C}^{(m)}$ depends on the voxel data x_m, and so these $L + 2$ numbers need to be stored and updated on a voxel-by-voxel basis.

The detailed computational steps are presented in Fig. 1. Algorithm 1a is used to initialize the independent-of-voxel data structure (first $L + 1$ rows of $\mathbf{C}^{(m+1)}$) when image $m + 1$ is available. Algorithm 1b is then applied to each voxel in image $m + 1$; it updates the last row of $\mathbf{C}^{(m+1)}$ and produces ρ. Before any voxel data are gathered, $\mathbf{C}^{(0)}$ is initialized to $\delta\mathbf{I}$, for some small positive δ (10^{-7} in our implementation). This will prevent division by zero in the first application of Algorithm 1a, which must otherwise be reformulated for the special case $\mathbf{C}^{(0)} = \mathbf{0}$. Because $\tilde{\mathbf{S}}^T\tilde{\mathbf{S}}$ is proportional to the covariance matrix of the current estimates of the amplitudes (α, γ_k) of the signal and detrending vectors, this initialization is equivalent to assuming a very large (δ^{-2}), but finite, variance for the initial estimate of α and γ_k in each voxel.

A fixed amount of computation per image is required, which works out to be $5L + 12$ operations per voxel to compute ρ; the storage space required is $L + 2$ locations per voxel. (Here, "operation" means a floating point addition, multiplication, or square root; "location" means a floating point value.) The computational cost compares favorably with a batch-oriented implementation of Eq. [2], assuming the vectors $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_L, \mathbf{r}\}$ are orthonormalized. Then $3L + 4$ operations per voxel, per image would be required to compute ρ. (This does not include the overhead of orthonormalization, which is directly analogous to the computation of the first $L + 1$ rows of $\mathbf{C}^{(m)}$.) We emphasize that this direct implementation of Eq. [2] does not lend itself to real-time computation, because it is nonrecursive. In our applications, L is small (2 or 3); by far the biggest computational loads are reconstruction and display.

STATISTICS OF ρ

The use of the correlation statistic ρ and its associated amplitude α for FMRI activation detection may be derived by modeling the data vector $\mathbf{x}$ as in Eq. [1] with the addition of white noise:

$$\mathbf{x} = \alpha\mathbf{r} + \sum_{k=1}^{L}\gamma_k\mathbf{s}_k + \epsilon\boldsymbol{\eta}. \qquad [4]$$

Here, $\{\alpha, \gamma_k, \epsilon\}$ are unknown parameters to be estimated from the measurements $\mathbf{x}$; the vectors $\mathbf{s}_k$, $\mathbf{r}$ are as before; $\boldsymbol{\eta}$ is a vector of independent $N(0, 1)$ (standard normal) random variables; ϵ^2 is the unknown variance of the noise in the chosen voxel (we do not assume that the noise has the same distribution in all voxels). In this section, α refers to the true (but unknown) value in Eq. [4] and $\hat{\alpha}$ refers to a statistical estimate from the data $\mathbf{x}$.

Under this model, the probability density for $\mathbf{x}$ is

$$P(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}\epsilon^N}\exp\left[-|\mathbf{x} - \alpha\mathbf{r} - \sum_{k=1}^{L}\gamma_k\mathbf{s}_k|^2/(2\epsilon^2)\right]. \qquad [5]$$

Maximum likelihood estimation (MLE) is a general technique for estimating unknown parameters (17). The MLE method is simply to maximize $P(\mathbf{x})$ in Eq. [5] by varying the parameters—the observed $\mathbf{x}$ is taken as fixed. For this

Algorithm 1a: Initialize for image $m + 1$

$\mathbf{z} \leftarrow [s_{1,m+1}\ s_{2,m+1}\ \cdots\ s_{L,m+1}\ r_{m+1}]$

$\beta_{\text{old}} \leftarrow 1$

for $j = 1, 2, \ldots, L + 1$ do

$h_j \leftarrow z_j/c_{jj}$ [save h_j]

$\beta_{\text{new}} \leftarrow \left(\beta_{\text{old}}^2 + h_j^2\right)^{1/2}$

$f_j \leftarrow \beta_{\text{new}}/\beta_{\text{old}}$ [save f_j]

$g_j \leftarrow h_j/(\beta_{\text{new}} \cdot \beta_{\text{old}})$ [save g_j]

$\beta_{\text{old}} \leftarrow \beta_{\text{new}}$ [save last β_{old}]

for $k = j, \ldots, L + 1$ do

$z_k \leftarrow z_k - h_j \cdot c_{kj}$

$c_{kj} \leftarrow f_j \cdot c_{kj} + g_j \cdot z_k$

Algorithm 1b: Update ρ, α in one voxel

$\hat{z} \leftarrow x_{m+1}$ [new MR intensity]

for $j = 1, \ldots, L + 1$ do

$\hat{z} \leftarrow \hat{z} - h_j \cdot c_{L+2,j}$

$c_{L+2,j} \leftarrow f_j \cdot c_{L+2,j} + g_j \cdot \hat{z}$

$c_{L+2,L+2} \leftarrow \left(c_{L+2,L+2}^2 + \hat{z}^2/\beta_{\text{old}}^2\right)^{1/2}$

$\rho^{(m+1)} \leftarrow c_{L+2,L+1}/\left(c_{L+2,L+2}^2 + c_{L+2,L+1}^2\right)^{1/2}$

$\alpha^{(m+1)} \leftarrow c_{L+2,L+1}/c_{L+1,L+1}$

FIG. 1. Computational recipes to update ρ in each voxel when one new image is calculated. Algorithm 1a is executed first; its inputs are the new components for the detrending vectors s_k and reference vector $\mathbf{r}$; its outputs are the coefficients $\{h_j, f_j, g_j, j = 1 \ldots L + 1\}$, and β_{old}. Algorithm 1b is then executed for each voxel; its inputs are the outputs of Algorithm 1a, and the new MR signal intensity for each voxel; its outputs are the updated values of ρ and α. Both algorithms also update the elements of the Cholesky factor $\mathbf{C}$.

linear, normally distributed model, MLE is equivalent to least squares estimation for $\{\alpha, \gamma_k\}$; the result for $\hat{\alpha}$ is given in Eq. [2]. The MLE for the noise variance is $\hat{\epsilon}^2 = N^{-1}|\mathbf{x} - \hat{\alpha}\mathbf{r} - \Sigma_k \hat{\gamma}_k \mathbf{s}_k|^2 = N^{-1}c^2_{L+2,L+2}$.

We may also compute the probability distribution of the estimated quantities. We find that $\hat{\alpha}$ is normally distributed with mean α and variance $\epsilon^2/|\mathbf{Pr}|^2$; we find that $\hat{\epsilon}^2$ is distributed like ϵ^2/N times a χ^2 variable with $\nu = N - L - 1$ degrees of freedom ($N - L - 1$ because $L + 1$ other parameters are being estimated from $\mathbf{x}$). We also find that the distributions of $\hat{\alpha}$ and $\hat{\epsilon}^2$ are independent. These facts together imply that

$$T = \frac{\hat{\alpha} - \alpha}{\hat{\epsilon}} \cdot |\mathbf{Pr}| \left(\frac{\nu}{N}\right)^{1/2} \qquad [6]$$

is distributed like a Student t variable with ν degrees of freedom.

We may use this result to test if $\hat{\alpha}$ is significantly different from zero. Under the null hypothesis ($\alpha = 0$), T in Eq. [6] is equal to $\nu^{1/2} c_{L+2,L+1}/c_{L+2,L+2}$, and so is readily computable from the algorithm presented in the Recursive Calculation of ρ section. Accepting as "activated" those voxels with $|T| > t_{\nu,p/2}$ will have a type I (false positive) error probability of p. Equivalently, a confidence interval of probability $1 - p$ for α is

$$|\alpha - \hat{\alpha}| < \nu^{-1/2} t_{\nu,p/2} \cdot \frac{c_{L+2,L+2}}{c_{L+1,L+1}}.$$

The threshold test on $|T|$ is the uniformly most powerful unbiased test, under the assumptions of the statistical model; the confidence interval for α is the shortest possible (17).

Under the null hypothesis, we find that ρ^2 is distributed like a beta variable with parameters ½ and ½ν. We may use this to develop a theoretical basis for choosing ρ_{thr}. (Thresholding $|\rho|$ is algebraically equivalent to using the t test on T.) Specifically, if we wish to accept a probability of false positives of p per voxel, then we choose ρ^2_{thr} to be the ordinate on the beta distribution with cumulative probability $1 - p$. Figure 2 gives an algorithm to approximate ρ_{thr}.

Algorithm 2: Compute ρ_{thr}

$$s \leftarrow \sqrt{-2\log(\tfrac{1}{2}p)}$$

$$x \leftarrow s - \frac{0.010328s^2 + 0.802853s + 2.515517}{0.001308s^3 + 0.189269s^2 + 1.432788s + 1} \qquad [26.2.23]$$

$$t_{\nu,p/2} \leftarrow x + \frac{x^3 + x}{4\nu} + \frac{5x^5 + 16x^3 + 3x}{96\nu^2} \qquad [26.7.5]$$

$$\rho_{\text{thr}} \leftarrow \frac{t_{\nu,p/2}}{\sqrt{\nu + t^2_{\nu,p/2}}} \qquad [26.5.27]$$

FIG. 2. Algorithm to approximate ρ_{thr}. Inputs are ν (nominally $N - L - 1$), and p = probability of false positive per voxel. Citations in square brackets are to equation numbers from ref. 23.

RESULTS

Echo planar data were acquired from a whole body Biospec 3 T scanner (Bruker Instruments), using a 30-cm inner diameter three-axis balanced-torque head gradient coil (18), and a quadrature transmit/receive endcapped birdcage RF coil designed for highly sensitive whole brain imaging (19). A single shot blipped gradient-echo EPI sequence using an initial 90° pulse and an effective *TE* (to reach $k_x = k_y = 0$) of 40 ms was used. The FOV was 24 cm, slice thickness 8 mm, and *TR* was 2 s (to ensure that no data was lost during real-time processing—the scanning system as described above can have *TR* as low as 100 ms). Image matrices of 64 × 64 and 96 × 96 have both been used; for display purposes, images presented herein were interpolated to 256 × 256.

The analog signal was diverted from the Bruker data acquisition system into a Crimson workstation (Silicon Graphics) containing a 14 bit A/D converter (Pentek). Real-time echo planar reconstruction was followed by the algorithms of Fig. 1. The brain images, with color overlays indicating voxels where $|\rho| \geq \rho_{\text{thr}}$, were displayed approximately 500 ms after each RF pulse.

Figure 3 shows some of the images resulting from our first real-time functional imaging experiment, conducted with a 64 × 64 image matrix, and using a repetitive *"finger-tapping:rest"* task paradigm (3). In this instance, the subject was instructed to tap the fingers of both hands sequentially for 20 s and then rest for 20 s, alternating for five full cycles (100 images). The reference time series $\mathbf{r}$ was the corresponding square wave with period 40 s, delayed 4 s from the task timing to allow for the hemodynamic rise time (7).

The background gray scale image in each panel of Fig. 3 is the echo planar image resulting from the first shot; it has a great deal of CSF-gray matter contrast and provides a good anatomical reference, with no need to warp the echo-planar derived overlays due to B_0 inhomogeneity induced image distortions. The brighter overlays in Fig. 3 were computed using a constant $\rho_{\text{thr}} = 0.45$, which is a value appropriate for the last image. An alternative display strategy would be to lower ρ_{thr} as m increases, so as to maintain the same p value. This is less interesting and less useful to observe in real time, because very few voxels become significantly activated until two task cycles are completed (at $m = 35$). Also, the "spuriously" correlated voxels convey some information. Although each individual correlation is not statistically significant for small m, the clustering of correlated voxels *is* significant. The large number of correlated voxels at the left edge of the brain at $m = 15$ and $m = 25$ are probably indicative of slight head motion at the start of the first finger tapping interval. The ability to monitor such effects is important in the applications of real-time FMRI.

FURTHER DIRECTIONS FOR REAL-TIME FMRI

As more computational power becomes attached to MR scanners, it will become routine to reconstruct and display multislice and 3D images in near real time. The work presented in this article shows that processing for FMRI activation may be done at small additional cost. An

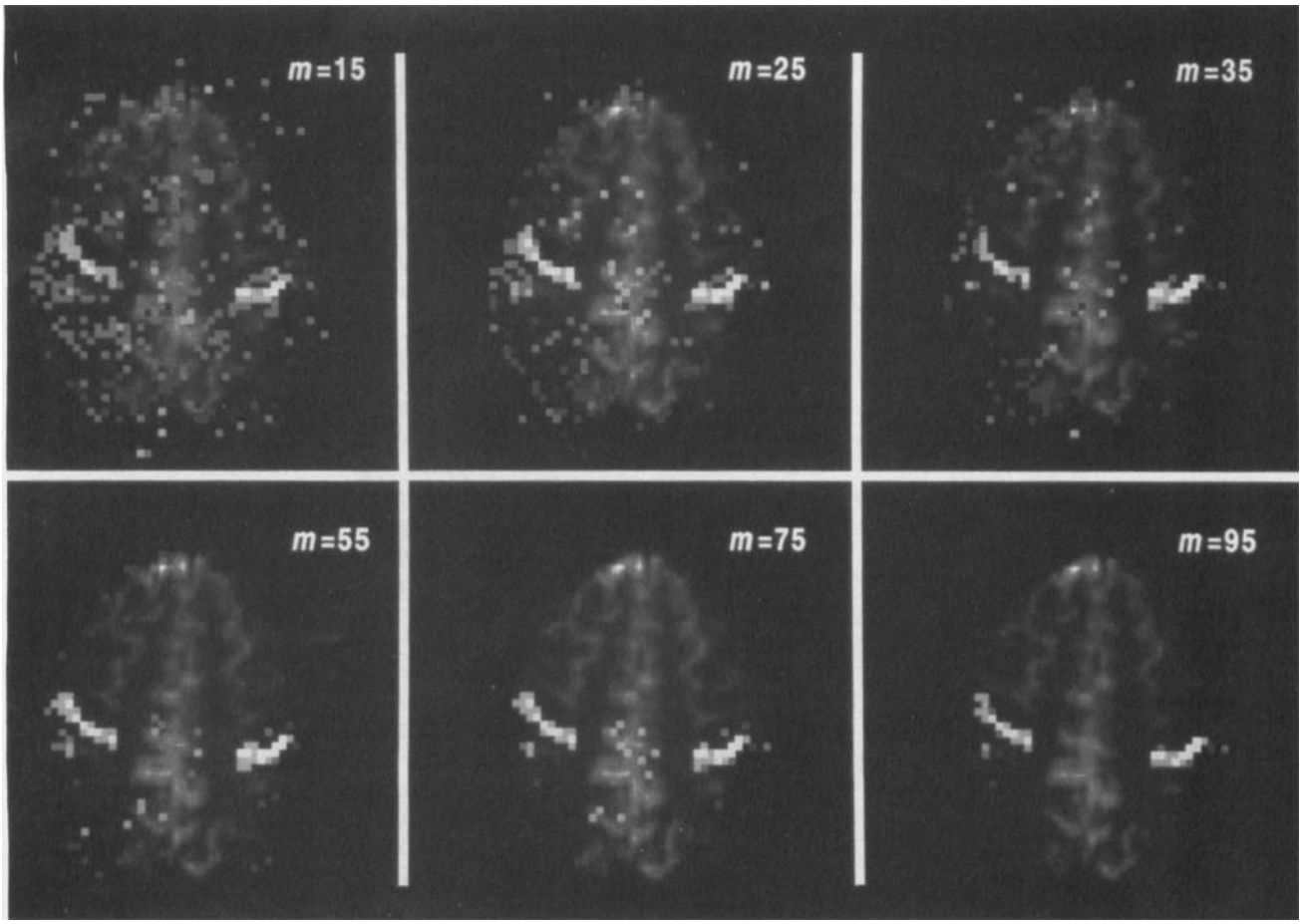

FIG. 3. Images from the first real-time FMRI experiment; task alternation was 10 images of sequential finger tapping, 10 images of rest; echo planar images are 64 × 64, FOV 24 cm. Brighter gray levels are overlaid on voxels with $\rho^{(m)} \geq 0.45$; white corresponds to the largest value of $\hat{\alpha}$. For all images, the background gray scale image is from the first echo planar shot, which has the greatest CSF-gray matter contrast. The gradual resolution of the subject's motor cortex is apparent.

important issue that remains to be creatively explored is the display of 3D functional results in real time.

Choice of the reference vector **r** is an issue addressed by Bandettini *et al.* (7), Friston *et al.* (20), and Binder *et al.* (21, 22). Uncertainty in **r** will lead to an increased probability of false negatives: to the extent that the model Eq. [4] is accurate, but **r** is incorrect, then true positives will tend to have $|\rho|$ underestimated and so be lost in the noise. One possibility not yet explored in the literature is to project each data vector **x** onto a multidimensional "response space" rather than a single vector **r**; this is the idea behind multiple correlation (12). An example would be to use two or three harmonics of a square wave (i.e., the Haar wavelets) for a periodic task alternation protocol, so that the reference function can be more flexible in its shape. After two or three task cycles, a "best fit" reference vector could be chosen from the response space, if the lower dimensional correlation is found to be more statistically robust. The recursive algorithm presented herein is readily adapted to computations of these sorts.

In principle, Algorithm 1a could be executed before data acquisition, using the detrending and reference vectors, and its results stored for each m. In practice, the amount of time saved would be very small. This means that it is practicable to compute real-time FMRI results with the detrending and/or reference vectors acquired in real time along with the MR images.

As pointed out by several investigators (8–10), the noise in FMR image sequences is not temporally white; that is, the density Eq. [5] should be modified to include correlations between the components of **x**. If this is done, the statistics of ρ become significantly more complicated; in general, we may expect correlated noise to cause larger values of $|\rho|$ to be more likely under the null hypothesis. This may be approximately allowed for by decreasing the number of degrees of freedom ν used to calculate the threshold. The simplest way to do this is to gather an image sequence of null data (no task alternations), and then compute the values of ρ in all brain voxels. Under the beta distribution, $\langle \rho^2 \rangle = (\nu + 1)^{-1}$; turning this around, we may estimate ν by $\hat{\nu} = (\overline{\rho^2})^{-1} - 1$. This provides a convenient way to approximate the distribution of ρ and extrapolate it (using Algorithm 2). The threshold ρ_{thr} may then be set corresponding to a lower p value than is directly available from the null data image sequence statistics.

It is common to choose quite small values of p (e.g., 10^{-4} or less) so as to ensure that the total probability of a false positive in an image is less than 0.05 (say). A simple

and conservative way to choose p is by the Bonferroni method: for example, if there are 1000 brain voxels in an image, then $p = 0.05/1000 = 5 \times 10^{-5}$ is the Bonferroni corrected threshold probability per voxel. Spatial correlations in the noise (as opposed to the temporal correlations discussed in the previous paragraph) mean that the Bonferroni corrected p can be very conservative. More elaborate methods for choosing p for FMR image analysis are discussed by Friston *et al.* (20).

APPENDIX: MATHEMATICAL NOTATION

The major symbols and notations used in this article are gathered here for ease of reference. A vector space formalism is used to represent signals as functions of time. For example, $\mathbf{x}$ is a vector whose mth component x_m is image intensity in a particular voxel at time $t = m \cdot TR$.

N = Dimension of data vectors = number of MR images.

L = Number of detrending vectors (i.e., time series to be removed from the data before the correlation coefficient is calculated).

$\mathbf{s}_k$ = Detrending vector k, for $k = 1, 2, \ldots, L$.

$\mathbf{r}$ = Reference vector (i.e., time series being correlated with the data). We implicitly assume that $\mathbf{r} \notin$ span $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_L\}$; that is, $\mathbf{Pr} \neq 0$ (see below for $\mathbf{P}$).

$\mathbf{x}$ = Data vector (i.e., time series of MR intensity *from a single voxel*).

α = Amplitude of $\mathbf{r}$ in $\mathbf{x}$, after detrending (constant that minimizes $|\mathbf{Px} - \alpha\mathbf{Pr}|$).

γ_k = Amplitude of $\mathbf{s}_k$ in $\mathbf{x}$.

ρ = Partial correlation coefficient of $\mathbf{x}$ with $\mathbf{r}$, when correlated components of detrending vectors are removed.

$|\mathbf{x}|$ = $[\mathbf{x}^T\mathbf{x}]^{1/2} = [\Sigma_j x_j^2]^{1/2}$ (N.B.: a weighted inner product $\Sigma_j w_j x_j y_j$ may be allowed for by scaling all vectors in $\mathbf{R}^N$ by $w_j^{1/2}$).

$\mathbf{S}$ = $N \times L$ matrix of detrending vectors: $[\mathbf{s}_1\ \mathbf{s}_2\ \ldots\ \mathbf{s}_L]$.

$\mathbf{P}$ = $N \times N$ projection matrix from $\mathbf{R}^N$ onto the orthogonal complement of span $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_L\}$.
= Operator that removes correlated components of detrending vectors from data vectors.
= $\mathbf{I} - \mathbf{S}[\mathbf{S}^T\mathbf{S}]^{-1}\mathbf{S}^T$ (N.B.: $\mathbf{P}^2 = \mathbf{P} = \mathbf{P}^T$).

$\tilde{\mathbf{S}}$ = $N \times (L + 2)$ matrix of detrending vectors, plus reference and data vectors: $[\mathbf{s}_1\ \mathbf{s}_2\ \ldots\ \mathbf{s}_L\ \mathbf{r}\ \mathbf{x}]$.

$\mathbf{C}$ = $(L + 2) \times (L + 2)$ lower triangular Cholesky factor of $\tilde{\mathbf{S}}^T\tilde{\mathbf{S}}$ (i.e., the matrix with $c_{ij} = 0$ for $j > i$, and such that $\mathbf{CC}^T = \tilde{\mathbf{S}}^T\tilde{\mathbf{S}}$).

$\mathbf{\Delta}$ = Lower right corner 2×2 submatrix of $[\tilde{\mathbf{S}}^T\tilde{\mathbf{S}}]^{-1}$.

m = Time index; quantities labeled with a superscript $^{(m)}$ are formed with the first m components of the relevant vectors; for example, $\tilde{\mathbf{S}}^{(m)}$ is an $m \times (L + 2)$ matrix which is formed from the first m rows of $\tilde{\mathbf{S}}$; quantities formed with a subscript m are formed using only the mth components of the vectors (e.g., $\boldsymbol{\sigma}_m$ below).

$\boldsymbol{\sigma}_m$ = $L + 2$ vector formed from the mth row of $\tilde{\mathbf{S}}$ (i.e., from information available at time $t = m \cdot TR$).
= $[s_{1,m}\ s_{2,m} \cdots s_{L,m} r_m\ x_m]^T$

ρ_{thr} = Threshold in the test $|\rho| \geq \rho_{\text{thr}}$, used to decide if a voxel time series is significantly correlated with $\mathbf{r}$.

ν = Number of degrees of freedom in the distribution of ρ, used to compute ρ_{thr}; nominally $N - L - 1$.

$\hat{\alpha}$ = Variables with a "hat" denote statistical estimates of parameters; see the Statistics of ρ section.

ϵ^2 = Variance of the noise in $\mathbf{x}$.

REFERENCES

1. J. W. Belliveau, D. N. Kennedy, R. C. McKinstry, B. R. Buchbinder, R. M. Weisskoff, M. S. Cohen, J. M. Vivea, T. J. Brady, B. R. Rosen, Functional mapping of the human visual cortex by magnetic resonance imaging. *Science* **254,** 716–719 (1991).
2. K. K. Kwong, J. W. Belliveau, D. A. Chesler, I. E. Goldberg, R. M. Weisskoff, B. P. Poncelet, D. N. Kennedy, B. E. Hoppel, M. S. Cohen, R. Turner, H. M. Cheng, T. J. Brady, B. R. Rosen, Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl. Acad. Sci.* **89,** 5675–5679 (1992).
3. P. A. Bandettini, E. C. Wong, R. S. Hinks, R. S. Tikofsky, J. S. Hyde, Time course EPI of human brain function during task activation. *Magn. Reson. Med.* **25,** 390–397 (1992).
4. S. Ogawa, D. W. Tank, R. Menon, J. M. Ellerman, S. Kim, H. Merkle, K. Ugurbil, Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proc. Natl. Acad. Sci.* **89,** 5951–5955 (1992).
5. S. Ogawa, T. M. Lee, A. R. Kay, D. W. Tank, Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci.* **87,** 9868–9872 (1990).
6. J. R. Binder, S. M. Rao, Human brain mapping with functional magnetic resonance imaging, *in* "Localization and Neuroimaging in Neuropsychology" (A. Kertesz, Ed.), pp. 185–212, Academic Press, New York, 1994.
7. P. A. Bandettini, A. Jesmanowicz, E. C. Wong, J. S. Hyde, Processing strategies for time-course data sets in functional MRI of the human brain. *Magn. Reson. Med.* **30,** 161–173 (1993).
8. B. Biswal, P. A. Bandettini, A. Jesmanowicz, J. S. Hyde, Time-frequency analysis of functional EPI time-course series, *in* "Proc., SMRM, 12th Annual Meeting, New York, 1993," p. 722.
9. R. M. Weisskoff, J. Baker, J. Belliveau, T. L. Davis, K. K. Kwong, M. S. Cohen, B. R. Rosen, Power spectrum analysis of functionally-weighted MR data: what's in the noise? *in* "Proc., SMRM, 12th Annual Meeting, New York, 1993," p. 7.
10. P. Jezzard, D. LeBihan, C. Cuenod, L. Pannier, A. Prinster, R. Turner, An investigation of the contribution of physiological noise in human functional MRI studies at 1.5 Tesla and 4 Tesla, *in* "Proc., SMRM, 12th Annual Meeting, New York, 1993," p. 1392.
11. J. V. Hajnal, R. Myers, A. Oatbridge, J. E. Schwieso, I. R. Young, G. M. Bydder, Artifacts due to stimulus correlated motion in functional imaging of the brain. *Magn. Reson. Med.* **31,** 283–291 (1994).
12. J. S. Bendat, A. G. Piersol, "Engineering Applications of Correlation and Spectral Analysis," Wiley, New York, 1993.
13. H. B. Keller, Numerical solution of bifurcation and nonlin-

ear eigenvalue problems, *in* "Applications of Bifurcation Theory" (P. H. Rabinowitz, Ed.), pp. 359–394, Academic Press, New York, 1977.
14. T. S. Chow, Matrices and linear algebra, *in* "Handbook of Applied Mathematics" (C. E. Pearson, Ed.), pp. 878–927, Van Nostrand Reinhold, New York, 1990.
15. P. S. Dwyer, The square root method and its use in correlation and regression. *J. Amer. Stat. Assoc.* **40**, 493–503 (1945).
16. N. A. Carlson, Fast triangular factorization of the square root filter. *AIAA Journal* **11**, 1259–1265 (1973).
17. G. Casella, R. L. Berger, "Statistical Inference," Wadsworth & Brooks/Cole, Pacific Grove, 1990.
18. E. C. Wong, P. A. Bandettini, J. S. Hyde, Echo-planar imaging of the human brain using a three axis local gradient coil, *in* "Book of Abstracts, 11th Annual Science Meeting, SMRM, 1992," p. 105.
19. E. C. Wong, G. Tan, J. S. Hyde, A quadrature transmit-receive endcapped birdcage coil for imaging of the human head at 125 MHz, *in* "Proc., SMRM, 12th Annual Meeting, New York, 1993," p. 1344.
20. K. J. Friston, P. Jezzard, R. Turner, Analysis of functional MRI time-series. *Human Brain Mapping* **1**, 153–171 (1994).
21. J. R. Binder, A. Jesmanowicz, S. M. Rao, P. A. Bandettini, T. A. Hammeke, J. S. Hyde, Analysis of phase differences in periodic functional MRI activation data, *in* "Proc., SMRM, 12th Annual Meeting, New York, 1993," p. 1383.
22. J. R. Binder, S. M. Rao, T. A. Hammeke, P. A. Bandettini, A. Jesmanowicz, J. A. Frost, E. C. Wong, V. M. Haughton, J. S. Hyde, Temporal characteristics of functional magnetic resonance signal change in lateral frontal and auditory cortex, *in* "Proc., SMRM, 12th Annual Meeting, New York, 1993," p. 5.
23. M. Abramowitz, I. A. Stegun (Eds.), "Handbook of Mathematical Functions," U.S. Dept. of Commerce, Washington, 1964.