Functional Magnetic Resonance Imaging in Real Time (FIRE): Sliding-Window Correlation Analysis and Reference-Vector Optimization

Daniel Gembris,[1,4]* John G. Taylor,[1,2] Stefan Schor,[1] Wolfgang Frings,[3] Dieter Suter,[4] and Stefan Posse[1]

New algorithms for correlation analysis are presented that allow the mapping of brain activity from functional MRI (fMRI) data in real time during the ongoing scan. They combine the computation of the correlation coefficients between measured fMRI time-series data and a reference vector with "detrending," a technique for the suppression of non-stimulus-related signal components, and the "sliding-window technique." Using this technique, which limits the correlation computation to the last *N* measurement time points, the sensitivity to changes in brain activity is maintained throughout the whole experiment. For increased sensitivity in activation detection a fast and robust optimization of the reference vector is proposed, which takes into account a realistic model of the hemodynamic response function to adapt the parameterized reference vector to the measured data. Based on the described correlation method, real-time fMRI experiments using visual stimulation paradigms have been performed successfully on a clinical MR scanner, which was linked to an external workstation for image analysis. Magn Reson Med 43:259–268, 2000. © 2000 Wiley-Liss, Inc.

Key words: Real-time fMRI; correlation analysis; sensitivity optimization; sliding-window technique

At a number of sites worldwide (1–7) real-time functional MRI (fMRI) has been developed to increase efficiency of fMRI experiments. It requires that the reconstruction of MR data sets, the image transfer to the processing unit, the calculation of functional images, and their display are accomplished in the time between successive acquisitions at the same slice location. Computer performance permitting, motion correction can be a further processing step. In Goddard et al. (6), where the results of using parallel computers for fMRI data-processing are discussed, the processing is referred to as "online analysis" (but not real-time analysis) in view of latency times in the order of minutes.

Real-time fMRI has the potential for improving efficiency in conceiving and performing neurophysiological and neuropsychological experiments. This improvement is achieved by an easier development of new paradigms or by supervising a subject's performance, e.g., by detecting motion artifacts that allows decisions during the study to restart a measurement being made. The clinical relevance of real-time fMRI arises from its potential for localizing important brain areas before or even during neurosurgical interventions. For a more detailed explanation of the advantages of real-time fMRI we refer the reader to (2).

One of the most commonly used methods for detecting neuronal activity is correlation analysis (8). This is performed by computing the correlation coefficients between the time series of the reference vector representing the expected hemodynamic response and the measurement vector of each voxel. A real-time correlation algorithm was proposed earlier by Cox et al. (7). This algorithm allows "detrending," the subtraction of non-stimulus-induced signal changes from the measured time series. It works in a cumulative manner, in other words the correlation coefficients refer to time series that are growing in length with each newly acquired MR data set of a measurement.

We present a comprehensive description of a new methodology for real-time fMRI and results obtained with its implementation in software. Based on the optimization of cost functions we have developed a more flexible real-time algorithm that combines the correlation analysis in a computationally efficient manner with detrending, the sliding-window technique, and with an optimization of the reference vector. With the sliding-window technique, it is possible to restrict the correlation computation to only the most recent data sets. Thus each new data set replaces the data set of the sliding-window buffer, which was acquired first. The advantage over the cumulative correlation analysis is the constant sensitivity of the correlation coefficients for changes between stimulation cycles in the signal response shape and amplitude. A distinct advantage of the sliding-window technique is its capability of quantifying physiological variability when combined with the reference-vector optimization. This method, which adapts the reference vector to the measurement data, increases functional sensitivity by taking into account changes in the signal response, which may occur due to noise, attention effects (9–11) or hardware instabilities.

THEORY

The advantage of correlation analysis over other methods for generating activation maps, such as the *t*-test (8,12) or the Kolmogorov-Smirnov (K-S) test (12), is that the reference vector can have arbitrary shape, which will be of primary importance for the determination of hemodynamic-response parameters.

[1]Institute of Medicine, Research Center Jülich GmbH, Jülich, Germany.
[2]Department of Mathematics, Kings College, London, UK.
[3]Central Institute for Applied Mathematics, Research Center Jülich GmbH, Jülich, Germany.
[4]Institute of Physics, University of Dortmund, Dortmund, Germany.
Grant sponsor: European Union (Biomed); Grant number: PL 950870.
Present address: Stefan Schor, Siemens Medical Systems, Erlangen, Germany.
*Correspondence to: Daniel Gembris, Institute of Medicine, Research Center Jülich GmH, Leo-Brandt-Str., D-52425 Jülich, Germany.
E-mail: D.Gembris@fz-juelich.de
Received 15 July 1998; revised 30 August 1999; accepted 29 September 1999.

Sliding-Window Correlation With Detrending

The definition of the correlation coefficient in the presence of detrending is (7,8):

$$\rho = \frac{\vec{x}_S \, \vec{r}_S}{|\vec{x}_S||\vec{r}_S|} \quad [1]$$

$$\vec{x}_S = \vec{x} - \sum_{i=0}^{L-1} \gamma_i \vec{s}_i \quad [2]$$

$$\vec{r}_S = \vec{r} - \sum_{i=0}^{L-1} \delta_i \vec{s}_i, \quad [3]$$

where $\vec{x}$ is the measurement vector of one voxel that is updated at every time step, $\vec{r}$ is the reference-vector, γ_i and δ_i, computed on a voxel-by-voxel basis, are the detrending coefficients belonging to the L detrending vectors $\vec{s}_i$. The detrended vector $\vec{x}_S$ and the vectors $\vec{s}_i$ are orthogonal as are $\vec{r}_S$, $\vec{s}_i$. For a faster computation of ρ, we use the expression

$$\rho = \frac{\vec{x} \, \vec{r}_S}{|\vec{x}_S||\vec{r}_S|}, \quad [4]$$

which is equivalent to Eq. [1] as can be shown by substituting Eq. [2] into Eq. [1].

The formulae for the detrending coefficients γ_i and δ_i are found by the minimization of the quadratic forms

$$|\vec{x} - S\vec{\gamma}|^2 \text{ and } |\vec{r} - S\vec{\delta}|^2. \quad [5]$$

with $\vec{\gamma} = (\gamma_1 \gamma_2 \ldots \gamma_L)^T$, $\vec{\delta} = (\delta_1 \delta_2 \ldots \delta_L)^T$ and the detrending matrix $S = (\vec{s}_0, \ldots, \vec{s}_{L-1})$ which contains the $\vec{s}_i$ as column vectors.[5] The reference vector is detrended in the same way as the measurement vector to account that not only apparent drifts, but also parts of the actual signal response are being removed.[6] The minimization leads to the equations (see Appendix A and (13))

$$(S^T S)\vec{\gamma} = S^T \vec{x} \quad [6]$$

$$(S^T S)\vec{\delta} = S^T \vec{r}, \quad [7]$$

where the vectors $\vec{\gamma}$ and $\vec{\delta}$ are then found by solving these sets of linear equations. Provided that the detrending vectors are the same for all voxels, this is done most efficiently by multiplication with the inverse matrix of $S^T S$, which only has to be calculated once for all pixels at one time step (we use the conjugate gradient method for that). Note that identical detrending vectors are sufficient to account for intervoxel differences in trends, since the weighting of the detrending vectors is calculated for each voxel individually.

The detrending vectors can either stem from predefined functions or measured data. In the latter case, the primary sources would be ECG or breathing sensors. For an a priori definition, sine functions and polynomials in which the ith polynomial component is increasing as t^i, are possible candidates. We have chosen the polynomials, which give, when combined with the "sliding-window technique," a suitable trend model.

The basic idea of this technique is to confine the correlation calculation to the N latest MR data sets. We cannot take standard methods in time-series analysis, such as Kalman-filter methods (14), because they are computationally inefficient for our problem, as would be exponentially weighting data inside our window (15). A Fourier transform-based sliding-window filtering technique, which was described in (16), is not applicable here, since, apart from the required computation time, we want to combine detrending (filtering) with real-time correlation analysis.

The sliding-window correlation is divided into two stages, the cumulative computation, when the window is being filled up with data, and the "steady state" computation. For both cases the computational load is constant and can also be independent of the window width, as we have shown. As a consequence of this efficiency, achieved by an iterative calculation of the correlation coefficient, the detrending vectors cannot be held constant, but each of them becomes time-dependent (except for the trivial case of a detrending vector with constant entries). Their elements have to be chosen as the elements of a "sliding-window" which is shifted over a series of numbers, the detrending series. The detrending coefficients γ_i and δ_i are now also time-dependent and have to be calculated anew for each position of the sliding-window. We now specifically consider the important case of detrending series chosen as single polynomial components. These series are growing monotonically in time, which leads to a global decrease of the detrending coefficients apart from a possible superposition of minor fluctuations. For a better interpretation of these coefficients, they can be linearly transformed to simulate the case of constant detrending vectors. If t_{sw} is the measurement time in the coordinate system of the sliding window, which has its origin at t_0, the corresponding transformation is given by the Taylor series:

$$\sum_{i=0}^{L-1} \gamma_i t^i = \sum_{i=0}^{L-1} \tilde{\gamma}_i (t_{sw})^i \quad [8]$$

where

$$\tilde{\gamma}_k = \sum_{i>k}^{L-1} \gamma_i \binom{i}{k} t_0^{(i-k)}. \quad [9]$$

[5]The global minimum of the first expression in Eq. [5], given by Eq. [2], corresponds to the component of the measurement vector $\vec{x}$ orthogonal to the detrending vectors. The minimum of the second quadratic term in Eq. [5] occurs when trying to "explain" the reference vector by a linear combination of detrending vectors.

[6]If a measurement vector is synthesized by adding an arbitrary linear-combination of detrending vectors to any chosen reference vector, the correlation coefficient recovers to 1 (up to numerical accuracy) after detrending.

Using this detrending approach we are ready to obtain an expression for ρ (this can be found by substituting Eq. [2] and Eq. [3] in Eq. [4]):

$$\rho = \frac{\vec{x}\vec{r} - \sum_i \delta_i \vec{x}\vec{s}_i}{\sqrt{h_1 h_2}} \quad [10]$$

$$h_1 = \vec{x}^2 - 2\sum_i \gamma_i \vec{x}\vec{s}_i + 2\sum_{i>j} \gamma_i\gamma_j \vec{s}_i\vec{s}_j + \sum_i \gamma_i^2 \vec{s}_i^2 \quad [11]$$

$$h_2 = \vec{r}^2 - 2\sum_i \delta_i \vec{r}\vec{s}_i + 2\sum_{i>j} \delta_i\delta_j \vec{s}_i\vec{s}_j + \sum_i \delta_i^2 \vec{s}_i^2 \quad [12]$$

The index variables *i* and *j* run from 1 . . . *L*, the number of detrending vectors. Subsidiary variables are introduced for each of the scalar products which allow the correlation coefficient to be calculated iteratively. In the initial accumulation phase, these subsidiary variables only have an additive contribution (from the most recent image data), as opposed to the stationary state where both additive and substractive contributions occur.

In the special case of baseline detrending ($L = 1$ and $s_1^T = (1\ 1\ .\ .\ .)$), for which the corresponding pseudo-code is given in Fig. 1, our iterative correlation computation outperforms a straight-forward computation of Eq. [1] already for $N = 4$.

```
j ← 0
For i = N..MAX_IMAGES do
  r_old ← r_j
  r_j ← ..          [new element of reference vector]
  b ← b − r_old + r_j
  d ← d − r_old^2 + r_j^2
  For k = 0, 1..#voxels − 1 do
    x_old ← x_{j,k}
    x_{j,k} ← ..      [new pixel]
    a_k ← a_k − x_old · r_old + x_{j,k} · r_j
    c_k ← c_k − x_old + x_{j,k}
    e_k ← e_k − x_old^2 + x_{j,k}^2
    h ← (N · d − b^2) · (N · e_k − c_k^2)
    ρ_k ← (N · a_k − b · c_k) / h^(1/2)
  j ← j + 1
  If(j = N) then j = 0
```

FIG. 1. Pseudo-code of the sliding-window correlation, consisting of the accumulation phase and the "steady-state" computation, which only considers baseline detrending. The code for the latter is directly shown. For the former case the underlined instructions or terms have to be modified: *N* has to be replaced by the current number of data sets, the underlined subtractive terms have to be removed as well as the whole last line and the commands for storing the oldest elements of the reference vector and the data set. Before the accumulation phase, all subsidiary variables have to be initialized at zero. This algorithm has been derived from Eqs. [6], [7] and [10]–[12] for one detrending vector with identical entries.

Statistical significance of the activation may be analyzed by testing either for $\rho = 0$ or $\alpha = 0$, as proposed in Cox et al. (7). Both tests are equivalent, but the latter can easily be extended to multiple regression.

Optimization of the Reference Vector

If optimization is employed the reference vectors become time dependent and the correlation coefficients are then most efficiently computed using Eq. [1] instead of Eq. [10], since a fully iterative update of the subsidiary variables for the evaluation of Eq. [10] is no longer feasible.

The neuronal activation, whose time course is assumed to be identical with the paradigm used (as in most studies), causes changes of the cerebral blood flow that become manifest in the fMRI signal. This signal is modeled by convolution of the paradigm with the hemodynamic response function, which should take into account time delay and smoothing of the blood flow regulation (17–20). A model function that takes into account a post-undershoot of the fMRI signal has been proposed in (21). To increase the correlation with the measured data, either the hemodynamic response function or the reference vector can be optimized with respect to parameters that describe these effects. Since the reference vector can always be generated by convolution of the paradigm with a response function, both approaches are mathematically equivalent. We use the former in keeping with previous work.

In post-processing it is common to use empirical knowledge for the choice of hemodynamic response parameters. Our method for reference-vector optimization allows the reference vector to be generated automatically, as required by real-time processing. Ideally, the optimization is performed for all brain voxels (identified by an intensity threshold criterion) separately, without using a priori information about the localization of activation. This approach is in contrast to most other related works, which in part assume a uniform response function over the entire brain (20). In practice, it may be advantageous to perform the optimization for groups of voxels to reduce computation time and to increase accuracy of parameter estimates. An alternative optimization approach presented recently (23) is not directly applicable to real-time fMRI due to its computational demands.

We require that the reference vector depends on a set of parameters $\{\omega_j\}$, such as the time of the gradual increase or decrease of blood flow, and optimize the following quadratic form with respect to the parameters ω_j, which can be done by gradient descent, or more rapidly by conjugate gradients. In addition to Eq. [5] we now have the further cost function

$$E = \left|\vec{x}_s - \alpha\vec{r}_S(\omega_j)\right|^2 \quad [13]$$

where the minimum value of *E* with respect to the linear parameter α occurs at the so-called 'overlap' value[7] (7)

[7] If one thinks of the reference and measurement vectors as steady functions (int) the optimal α would maximize the geometrical overlap between them.

$$\alpha = (\vec{x}_S\vec{r}_S)/\vec{r}_S^2 = \rho|\vec{x}_S|/|\vec{r}_S|. \quad [14]$$

This relation allows α to be calculated. The appropriate statistical test for activation detection using nonlinear regression has been described in (23). It is shown there that for a sufficiently large sliding-window length, this test is well approximated by the *F*-test (12).

In our implementation of the optimization, detrending only consists of substracting a baseline that varies as the window moves. For the hemodynamic response function we have chosen the simplest asymmetric function possible:

$$h(t) = (t - b)e^{-(t-b)/a}\theta(t - b) \quad [15]$$

(where $\theta(x)$, the Heaviside θ function, is defined as $\theta(x < 0) = 0$, $\theta(x > 0) = 1$ and $\theta(x = 0) = \frac{1}{2}$) that we call "linear-exponential." This function, which is depicted in Fig. 2 together with the Gaussian function used by other researchers (18), has the useful property that the convolution with an on-off paradigm can be performed analytically. There are only two free parameters (a, b) to assure that, even for a short window width, the adaptation to those time-series that do not show stimulus-related activation is limited. The first parameter, a, gives the decay and the second, b, the extra delay time. Both parameters together specify the position of the maximum of $h(t)$, which is at $b + a$. The on-off box-car design function is convolved with $h(t)$ giving the reference function $r(t)$ which is inserted into Eq. [13]. The design function can in general consist of an arbitrary number of activation cycles, but for our work we restricted it, for the ease of programming, to only a single on-period followed by a single off-period. The convolution results in a nonlinear two-dimensional (2D)-optimization problem in the parameters a and b, which is solved by the method of conjugate gradients (12,24) (see Appendix B for details).

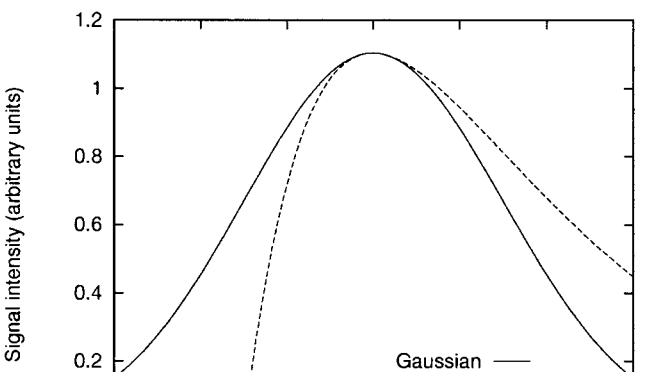

FIG. 2. Comparison of two different hemodynamic response functions with similar dispersions. The linear-exponential function (dotted line), which we introduced mainly for mathematical reasons (see Optimization of the Reference Vector), has typically a sharper signal increase and a flatter decrease than the Gaussian function (solid line). This asymmetry of the linear-exponential model is useful for modeling in vivo fMRI data. By normalization, both functions coincide near the maximum in this plot (a = 3 sec, b = 3 sec, ς = 3 sec), i.e., the values and the second derivatives at the maximum are the same for both functions.

As the sliding-window moves there are two partial activation cycles within the sliding-window buffer. In order to gain statistical power, we reduced the number of free parameters by optimizing only one response function, i.e., one delay and one dispersion parameter, for the sliding-window time-series data of one voxel. In our implementation, this is very conveniently achieved by performing an optimization after a reordering of the data, such that the contents of the sliding window is rotated backwards by t_i modulo N elements, where t_i is the current measurement time point ($t_i = 0$ for the beginning of the measurement), which was realised by a cyclic overwriting of the sliding-window buffer. This leads to a transition behavior of the response parameters when response characteristics change between activation cycles.

The optimization parameters are initialized with constants at the beginning.[8] After that the calculated parameters are fed back for the optimization at the next time step.

EXPERIMENTAL SETTING AND METHODS

Measurements were performed using a 1.5 T whole body scanner (Magnetom Vision, Siemens Medical Systems, Erlangen, Germany) with EPI booster. In our implementation of real-time fMRI, based on the software package FIRE (functional imaging in real time) developed in-house (25–27), the MR scanner itself reconstructs the images in real-time which are then transferred from the host computer of the scanner (Sun Sparc) via a TCP/IP connection (Ethernet) to an external workstation, a SUN Ultra Sparc 1 (143 or 170 MHz). FIRE has two main constituents, the real-time server, located on the host computer of the scanner, and the client, which is located on the workstation. The data processing, mainly consisting of correlation analysis and image interpolation, is performed by the client. Color-encoded values of the correlation coefficient are only displayed for those voxels whose image intensity exceeds 15% of the average image intensity.

Twenty healthy male volunteers with an age range of 20 to 35 years have been investigated in our studies. They gave written, informed consent for participation in the study, which was approved by the University of Düsseldorf Human Subjects Review Committee. The subjects have been visually stimulated using red flickering light with a frequency of 8 Hz. This was analyzed in two separate studies.

[8]Alternatively approximation formulae can be used to guide and thus to speed up the optimization computation. In the (practically relevant) case that the sliding-window comprises exactly one activation cycle, such formulae could be obtained by equating the analytical expressions for the momenta of the reference function with the corresponding momenta of the measured activation time course. This will in general lead to nonlinear systems of equations, but in special cases analytical solutions exist. Using only the first momentum, which corresponds to calculating the "center of mass" of the signal in the time domain, an approximation for the delay of the signal can be found.

For the first study, we defined 16 parallel slices with a matrix size of 32×32, which were measured using an EPI pulse sequence (TR = 2.2 sec, TE = 65 msec, voxel size: $6 \times 6 \times 6$ mm^3). The stimulation paradigm consisted of five baseline measurements and repeated alternation between five measurements for the active and five for the passive condition (each measurement covering all 16 slices). Detrending was performed with one vector of identical entries (substraction of the mean as in conventional correlation); the width of the sliding-window was 10 measurements (i.e., one cycle). The reference vector $\vec{r}$ has been chosen as the convolution of a Poisson-type hemodynamic-response function with the paradigm (the linear-exponential function was used in the project later). The parameter λ of the Poisson function, which specifies its width and delay simultaneously, was 3 sec. We extended the discrete function given in Friston et al. (17) to a continuous one and introduced a second parameter, analogous to parameter b in Eq. [15], for specifying an additional delay time, chosen to be 4 sec in this study. The choice of the parameter values (3 sec resp. 4 sec) appeared appropriate for our data, but its optimality has not been verified.

In the second study, we applied the reference vector optimization described in Optimization of the Reference Vector to data from a different, previously conducted visual stimulation experiment. (EPI, effective TR: 1 sec, TE: 67 msec, number of measurements: 512, matrix size: 32×32, voxel size: $6 \times 6 \times 6$ mm^3, 12 initial baseline measurements, 10 stimulation cycles with 10 activation and 40 baseline measurements). To increase signal-to-noise ratio (SNR), time series were generated by averaging over nine neighboring voxels (from the visual cortex). The timing of the visual stimulation was synchronized to the pulse-sequence trigger of the scanner. The optimization of the reference vector, which at the time when our study was conducted had not yet been part of FIRE, has been performed offline on different computer systems (LINUX-PC, Sun Ultra Sparc 1 and Cray T3E) using our SEPP (Single Event Processing Package) software.

RESULTS AND DISCUSSION

Sliding-Window Correlation With Fixed Reference Vector and Detrending

A typical correlation map of the first study, obtained in a single activation cycle lasting 22 sec, is shown in Fig. 3. In the visual cortex, the correlation values exceed the thresh-

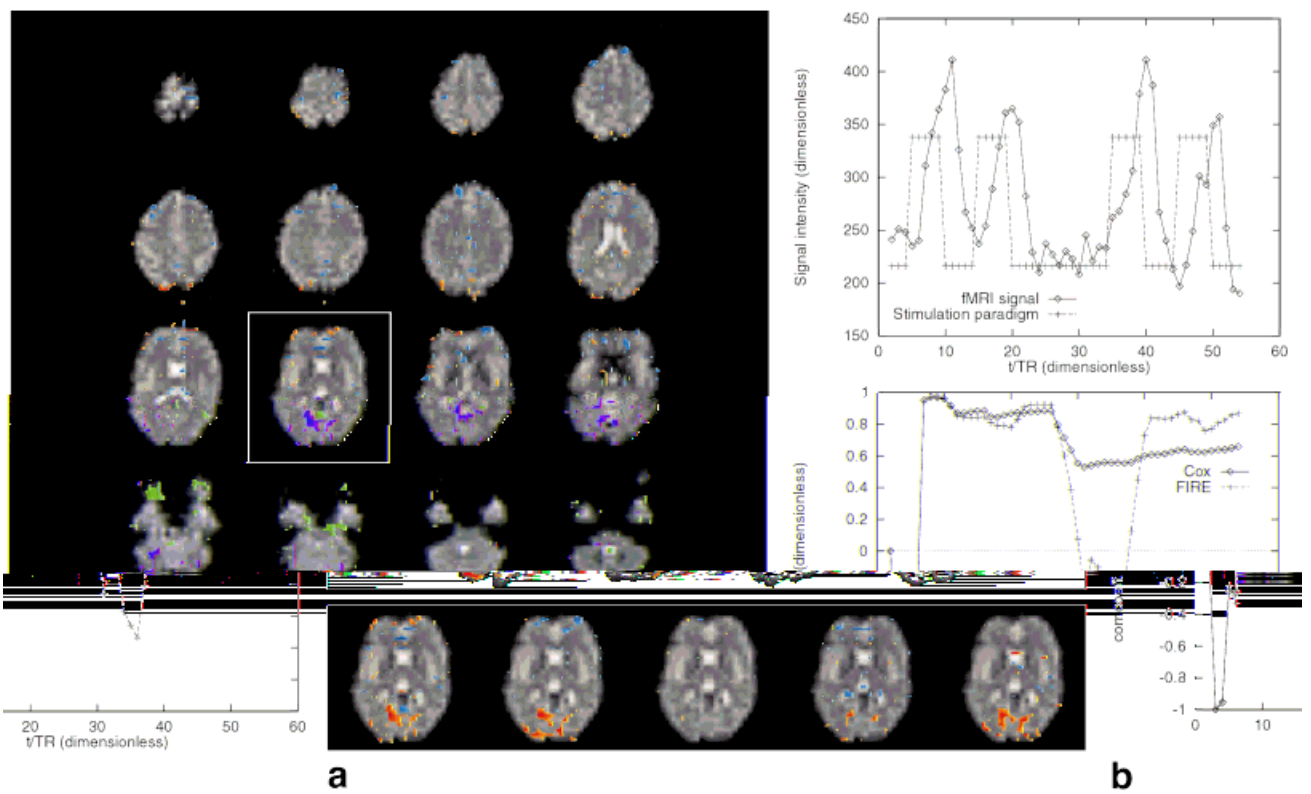

FIG. 3. The influence of a missing stimulation phase on the activation detection in a periodic paradigm. **a:** Brain activation (color overlay over raw images) during a visual stimulation study, where one stimulation of the repetitive paradigm has been omitted. Despite the short width of the sliding-window, which covers a single activation cycle, activation in the primary visual cortex is clearly visible. The activation maps for the framed slice and those time points for which the sliding-window covers exactly one activation cycle are shown at the bottom. **b:** Time-courses of the paradigm, the fMRI-signal (for a region-of-interest in the visual cortex), and the correlation coefficient obtained with both the sliding-window computation and the cumulative correlation computation. The change in activation is more distinct in the case of the sliding-window correlation. This method also shows a faster recovery of the correlation coefficient to the value obtained before the stimulus omission (all plots in this paper were generated using GNUPLOT, version 3.5).

old of 0.70, which indicates a good correspondence between the measured and the model time series. Signal changes in the visual cortex could be clearly detected in a single activation cycle in similar studies conducted on the other 19 subjects.

In principle, the only limitation of L, the number detrending vectors, is that it must always be smaller than the length N of the sliding-window, $L < N$. In practice $L \ll N$ has to be fulfilled in order to restrict detrending of activation related signal changes. For a single activation cycle per sliding-window, we noticed that the correlation coefficient already decreased for second-order detrending, corresponding to a correction of baseline shifts and a removal of linear trends, which confirms findings in (28). The main conclusion of that work is that it is beneficial to design paradigms that are symmetric about the midpoint of the time series. We found that the observed decrease in the correlation coefficient depends on how well the reference vector can be described by a superposition of a given number of detrending vectors, which is determined by the type of response function and the number of activation cycles in a sliding-window with a given length. The decrease of the correlation coefficient becomes less pronounced with an increasing number of activation cycles, since less signal is attributed to trends. It has to be distinguished from a decrease of the correlation coefficient with an increase in the order of detrending. This corresponds to a decrease in statistical significance expressed by a p value, which is related to ρ by a transformation given in (7).

In general, if regressors are not orthogonal, then detrending is conservative, since the signal fluctuations in the data that are assigned to the discarded detrending space are assumed to arise from *nonsignal* sources (i.e., "undesired" signal components). Only if some a priori information about the detrending space is available would it be possible to reassign some of this variation to the signal (e.g., if one knew that the linear slope of the data due to nonsignal sources was bounded, then a slope that is larger than the upper limit could be inferred to be from a signal).

The observed reduction of the correlation coefficient implies that detrending beyond a baseline correction, as described in the theory section and implemented in our software, should be applied in a particularly careful manner to time series consisting of only one activation cycle, such as those obtained by single-trial activation studies.

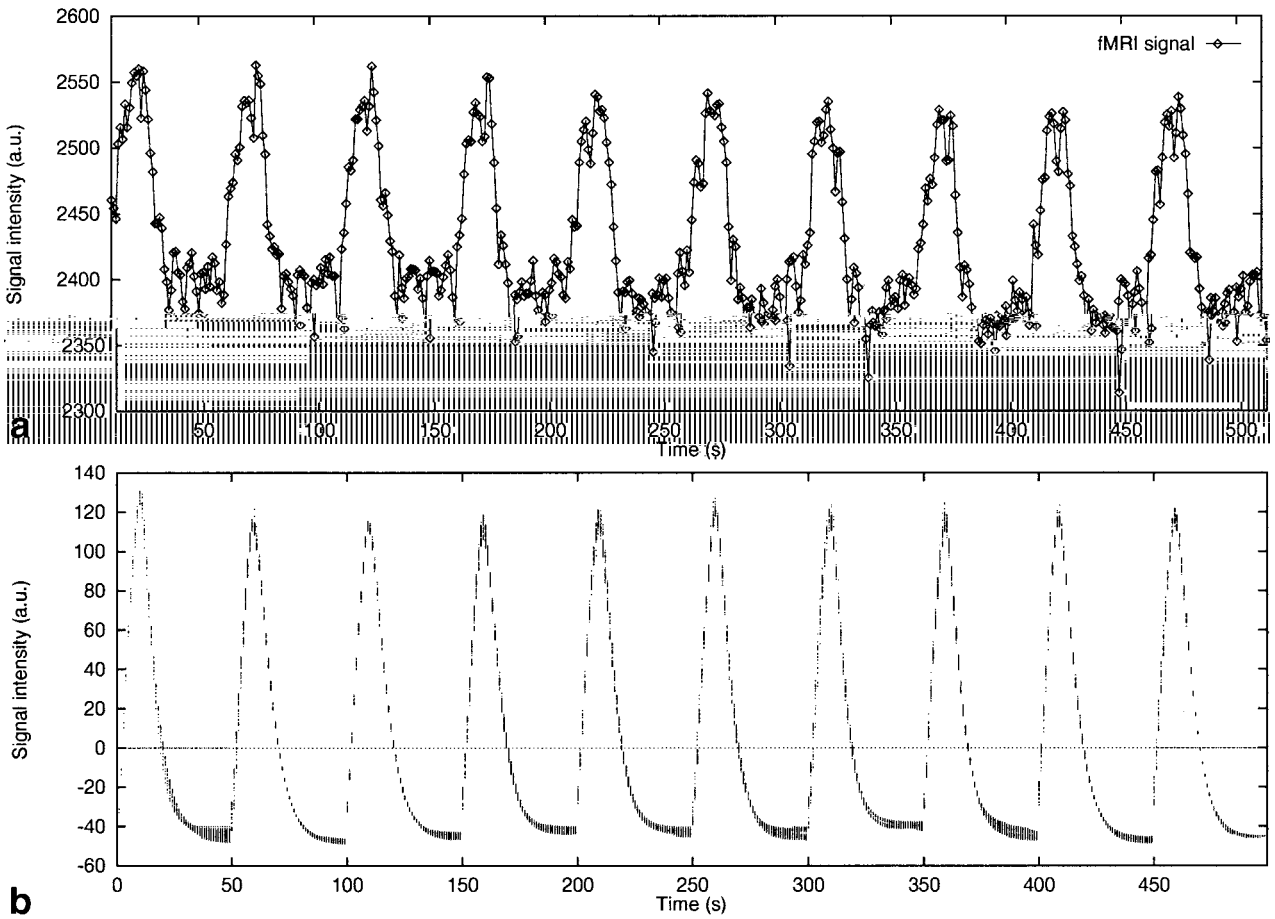

FIG. 4. **a:** Time-series from a region of interest in the visual cortex comprising 3 × 3 neighbouring voxels (voxel size: 6 × 6 × 3 mm^3) obtained with a repetitive visual stimulation paradigm. **b:** Superposition of all 500 reference vectors, each of which is 50 time points long, displayed as a scatter plot. These vectors have been obtained for the above time-series by the optimization of the reference vector using the linear-exponential model for the hemodynamic response function. The plot demonstrates the good stability of the optimization routine and reveals slight variability in the shapes of the optimal reference vectors.

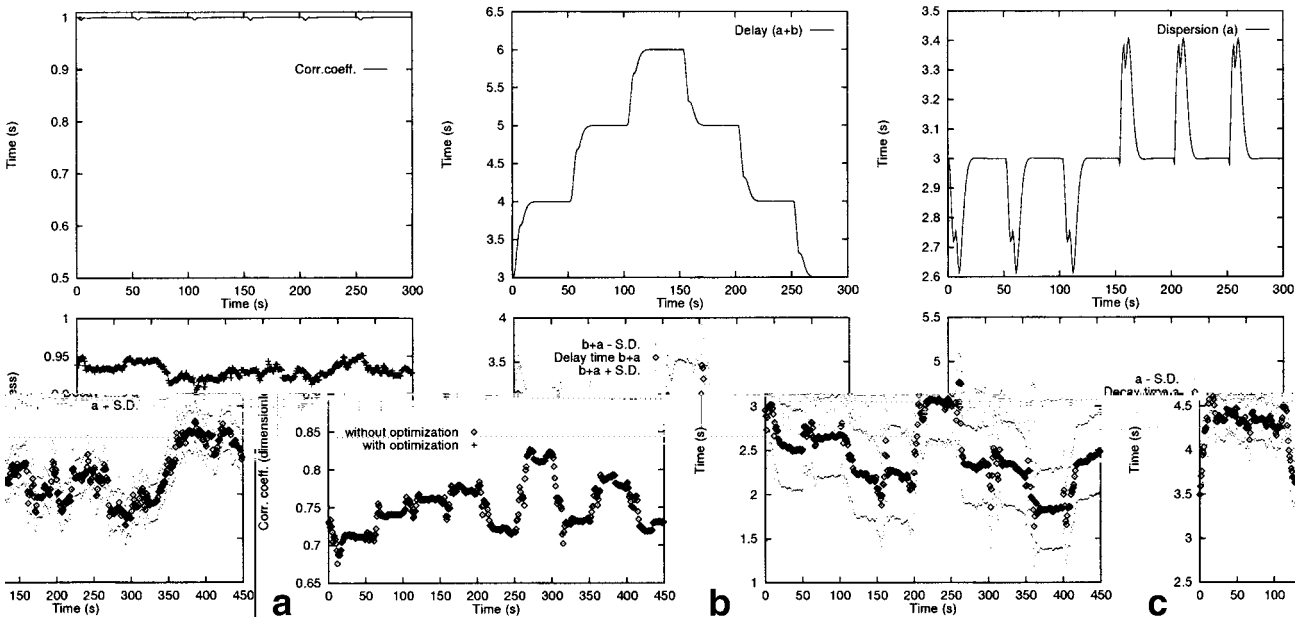

FIG. 5. Time-courses for the correlation coefficient ρ (**a**), hemodynamic delay (**b**) and dispersion (**c**) obtained by the optimization of the reference vector for synthetic (top) and experimental data (bottom). The synthetic data was generated by convolution of the paradigm belonging to Fig. 4 using the same linear-exponential response function (Fig. 2) that is employed for modeling the experimental data, where the delay time from one "activation cycle" to the next has been changed by 1 sec. This varying delay is reconstructed correctly by the optimization algorithm. In the intermediate stages, when the sliding-window buffer contains data of two consecutive activation cycles, the model function cannot describe the data in the buffer perfectly, which results in a slight decrease of the correlation coefficient to 0.99 and the deviation of the dispersion parameter *a* from the correct value. The time courses in the lower row have been created for the experimental data depicted in Fig. 4. The reference vector for each time point was modeled by convolution of the stimulation paradigm for the previous 50 time points with the linear-exponential response function. Therefore the range of the *t* axis comprises only 450 instead of 500 time points. The standard deviation of the delay and decay times, obtained by a Monte Carlo simulation, are indicated by the thin dotted curves. Note the increase in the stability of the correlation coefficient obtained with optimization as compared to a correlation computation with an optimally time-shifted box-car function. See Sliding-Window Correlation with Iterative Optimization of Reference Vector for a discussion of the increase in the amplitude. The time courses of the parameters show a transition behavior as discussed in the theory section.

Sliding-Window Correlation With Iterative Optimization of Reference Vector

Figure 4 depicts an example of a time series from a visual stimulation study that was generated by averaging over 3 $\times$ 3 voxels in the visual cortex.

In Fig. 5, the time courses of the decay time a and the delay time $b + a$ of the hemodynamic response function obtained by iterative optimization are shown, both for a simulated data set and the in vivo data shown in Fig. 4. The accuracy of our results has been assessed with a Monte-Carlo simulation (12), for which the fitting process has been repeated 100 times with Gaussian distributed random numbers, with a standard deviation $\varsigma = 25$, added to the measurement data shown in Fig. 4, from which the standard deviation ς of the noise was determined. Comparing the average standard deviation for the calculated parameters one finds that changes in the decay parameter can be detected with a greater sensitivity than in the delay parameter (0.3 sec versus 0.5 sec). Figure 5 also depicts the time courses of the correlation coefficient obtained with and without reference-vector optimization. The mean and standard deviation of the correlation coefficient was found to be 0.93 $\pm$ 0.01 and 0.75 $\pm$ 0.03, respectively.

In the case of no optimization, a box-car reference function with a delay of 4 sec is used. Of all integer numbers smaller than 8, this delay time resulted in the highest correlation value. The box-car function, which is not very well suited for the measured data, has been chosen for its simplicity, since it only has one free parameter, the delay time. For this we choose 4 sec, as of all integer numbers smaller than 8 it resulted in the highest correlation value.

These numbers have to be compared with 0.064 $\pm$ 0.125, the mean and standard deviation of ρ found by optimization on the time course of a baseline study. For such time courses, the optimization takes longer, which makes it advisable, for the processing of two- or three-dimensional (2D or 3D) data sets, to apply a simple test beforehand to avoid reference vector optimization for nonactivated voxels. Already one or two conjugate-gradient steps are enough for increasing the correlation coefficient close to final value. Further iterations are needed for the parameter values to settle (typically about five in total).

The plot in the lower part of Fig. 4 depicts the superposition of all optimized reference vectors calculated iteratively at different time points, which can be interpreted as a "filtered" brain response signal. The superposition shows a good coherence of the reference vectors for subsequent sliding-window positions. This plot demonstrates the stability of our algorithm. To assess the model dependence of the parameter changes, we repeated the computations using a trapezoid function for the reference-vector realized by the product of two tanh functions. The param-

eter time courses resemble largely the previous ones. The apparent variability in a and b found in vivo is caused is in part due to changes in the shape of the measured signals over time, which are not considered in our linear-exponential model. For more noisy activation data (SNR = 1:1) fitting was still possible, but the computed parameters then showed larger fluctuations over time. To elucidate the physiological meaning of the hemodynamic response parameters, e.g., their relevance for "mental chronometry," the measurement of quantities independent of fMRI would be required, as for example response times (29).

For the LINUX-implementation of the reference-vector optimization the average computation time per sliding-window position was 0.49 msec on a Pentium II-400, allowing real-time processing of 3D fMRI data sets. If the parameter feedback is turned off, the computation time increases by about one order of magnitude. A computationally more demanding algorithm, the uniform gridding of parameter space for the cost function Eq. [13] and the linear response function described in Optimization of the Reference Vector with 0.25 sec resolution and 3,200 grid points, which has been implemented as part of the FIRE version for the Cray-T3E supercomputer (we used 64 of the 256 processing element, alpha processors clocked with 450 MHz) showed similar performance as the conjugate-gradients based reference-vector optimization on the PII-400.

CONCLUSION

In this work, we have presented an algorithm to compute correlation coefficients between fMRI time-series and reference time-series, which is suitable for (but not restricted to) real-time processing. It combines a voxel-by-voxel detrending, the sliding-window technique, and a reference-vector optimization in a computationally efficient manner. The sliding-window technique provides constant sensitivity to changes in functional activation during the entire scan. The reference-vector optimization, which allows the on-line computation of physiological parameters, such as delay and dispersion of the hemodynamic response function, and also the possibility of quickly changing a paradigm (over a time period corresponding to the length of the sliding-window) during a measurement, differentiates our dynamic approach from the cumulative approach by Cox et al. (7). Since the optimization is computationally the most demanding processing step, our real-time software (FIRE) (25–27) allows to deactivate this function, which makes it suitable also for standard UNIX workstations. Using FIRE with disabled optimization, functional maps of a single cycle of visual activation have been calculated and displayed in real-time with a sliding-window width that corresponds to 20 sec of scanning time.

It has been shown that today's PC technology has the potential for performing the reference-vector optimization for a sliding-window on 3D fMRI data sets. A further increase in computer performance can be used for improved discrimination of local and global minima in the optimization process. The possibility of detecting shifts of the delay time of the hemodynamic response opens up exciting opportunities for real-time studies targeting the investigation of higher cognitive functions (30). Currently studies are in progress, that use the methodology described in this paper for biofeedback experiments (31).

The observed decrease of the correlation coefficient using detrending could be studied further using a different, unified mathematical approach. This should treat the reference vector optimization and detrending, which have their roots in the minimization of the costfunctions Eq. [5] and Eq. [13], as a single nonlinear optimization problem in the hemodynamic parameters *and* the detrending coefficients.

ACKNOWLEDGMENTS

We thank Michael Peyerl, Randall Kroeker and Edgar Mueller (Siemens Medical Systems, Erlangen) for their

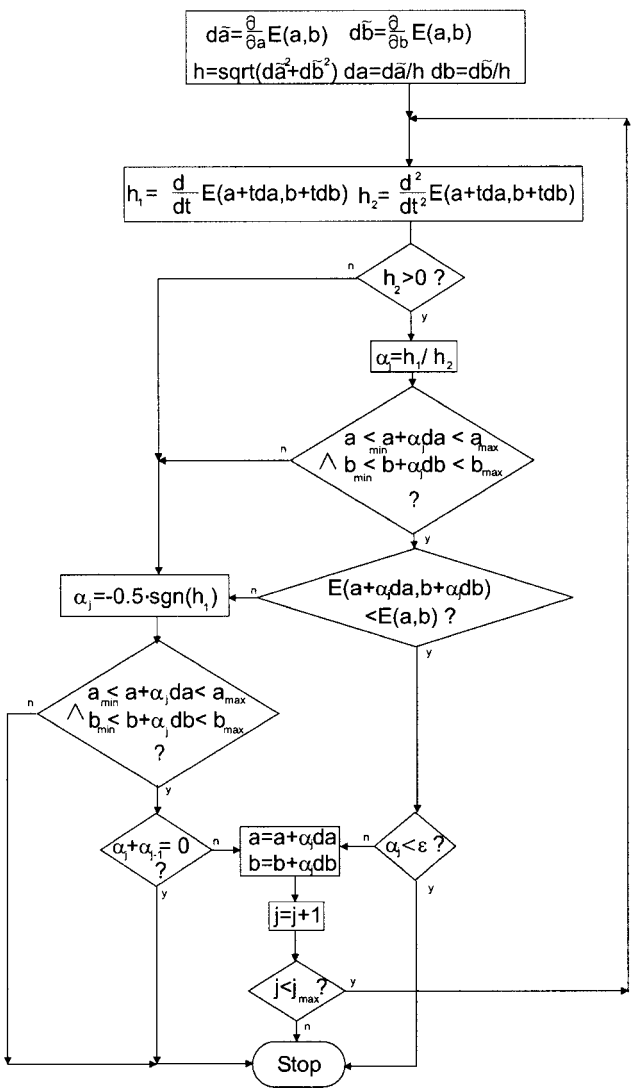

FIG. 6. Flow-chart of the ECO (enforced-convergence optimization) line-search algorithm used for our implementation of the method of conjugated gradients. If the line search is stopped, either the calculated or default parameter values are returned.

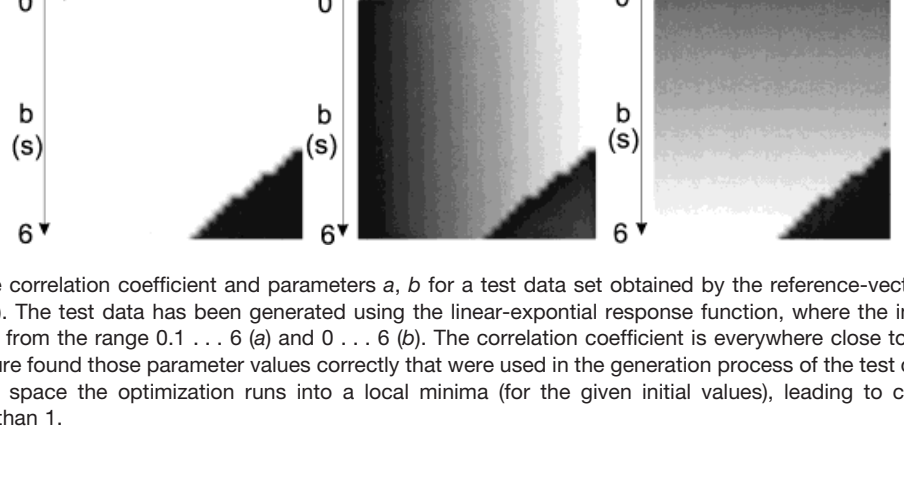

FIG. 7. Maps of the correlation coefficient and parameters *a*, *b* for a test data set obtained by the reference-vector optimization (initial values: $a = 3$, $b = 0$). The test data has been generated using the linear-expontial response function, where the input parameters were equidistantly chosen from the range 0.1 . . . 6 (*a*) and 0 . . . 6 (*b*). The correlation coefficient is everywhere close to 1, indicating that the optimization procedure found those parameter values correctly that were used in the generation process of the test data. In the lower right corner of parameter space the optimization runs into a local minima (for the given initial values), leading to correlation coefficients significantly smaller than 1.

help with modifying the scanner operating system for the FIRE project.[9] Valery Kiselev (Research Center Jülich) has made several valuable suggestions, especially concerning the optimization of the reference vector. We are also grateful for the support of the following members of our MR group: Maria-Liisa Grosse-Ruyken, Stefan Wiese, Barbara Elghahwagi, Michael Daunderer and Klaus Mathiak. One of us (J.G.T.) would like to thank the past and present Directors of IME, Prof. H.W. Müller-Gärtner and Prof. K. Zilles, for hospitality at IME during the performance of this work. Jon Shah and Bernhard Steffen (both Research Center Jülich) gave useful suggestions about the manuscript during the review process.

APPENDIX A: DETRENDING

We give here the derivation of the formulae [6] and [7] stated in Sliding-Window Correlation with Detrending.

Since

$$|\vec{x} - S\vec{\gamma}|^2 = 2(\vec{\gamma}^T S^T S\vec{\gamma} - \vec{x}^T S\vec{\gamma} + \vec{x}^T\vec{x})$$

then optimizing $|\vec{x} - S\vec{\gamma}|^2$ is equivalent to finding the minimum of the following quadratic form:

$$f(\vec{\gamma}) = \vec{\gamma}^T A\vec{\gamma} - \vec{b}^T\vec{\gamma} + c$$

$$A = S^T S$$

$$\vec{b} = S^T\vec{x}$$

$$c = \vec{x}^T\vec{x}$$

The minimum is obtained by solving the linear equations:

$$A\vec{\gamma} = \vec{b} \qquad [A1]$$

Minimization of $|\vec{r} - S\vec{\gamma}|^2$ gives an analogous expression for $\vec{\delta}$:

$$A\vec{\gamma} = S^T\vec{r} \qquad [A2]$$

Equations [A1] and [A2] are the same as [6] and [7] in the text.

APPENDIX B: OPTIMIZATION ALGORITHM BASED ON THE METHOD OF CONJUGATE GRADIENTS

For finding the minima of the cost function Eq. [13] the method of conjugate gradients (12,24) is used. This method reduces an *n*-dimensional optimization problem to minimizing a cost function along straight lines through search space, to so-called "line searches." They require the calculation of the first- and second-order derivatives of Eq. [14] along the search direction. Currently they are calculated numerically by difference quotients.

The development of a line-search algorithm which works fast and also efficiently for time series which do not match the response model (e.g., in nonactivated voxels) turned out to be one of the major tasks of our project. In our line search, termed ECO (enforced-convergence optimization), as in the established Marquard's (12) method both gradient descent and Newton's method are used. For Newton's method, which has extremely fast convergence near a minimum ("quadratic convergence"), the step size for the line search is calculated by dividing the first-order by the second-order derivative at the same point. Small values for the second derivative are thus a very "powerful" source of numerical instabilities. ECO avoids them by restricting the optimization to a pre-defined parameter range ($a \in [0.1 \ldots 10]$ and $b \in [-10 \ldots .10]$) and by explicit evaluation of the costfunction to exclude those Newton steps that would lead to a deterioration of the optimization result compared to the current result. The Newton iteration stops when a certain accuracy (here 0.001 sec) is

[9]Further information about FIRE is available on our WWW server http://ime-web.ime.kfa-juelich.de/ime_www/nmr/english/FIRE.html.

reached or when a maximum number of iteractions has been exceeded. For the *two*-dimensional case the flowchart of ECO is depicted in Fig. 6; it can easily be generalized to higher dimensional optimization problems.

In the gradient descent case, steps of uniform size, which have been adapted to the given problem (0.5 sec is adequate for our optimization), are performed.

The method has been verified as a part of reference-vector optimization (see Fig. 7). Although the optimization does not necessarily converge to the nearest minimum for given initial values, this is no guarantee in general that the global minimum has been found. One way of handling local minima, which become more frequent if the model function does not match well with the experimental data, e.g., if they are too noisy, is to perform the optimization for different initial values "in parallel" (the method of multiple restarts). Recently, robust methods for global optimization have been developed, e.g., Gorse et al. (32) and Barhen et al. (33). Instead of calculating the derivatives of Eq. [13] numerically, it is feasible to find analytical expressions for the them (which are most effectively obtained by the help of computer-algebra systems, such as MAPLE, which can directly translate them in to C source-code). This would reduce the computation time significantly.

REFERENCES

1. Hong X, Cohen M, Roemer P. Functional EPI with real time imaging processing. In: Proceeding, ISMRM, 5th Annual Meeting, Vancouver, 1997. p 321.
2. Voyvodic JT. Real-time fMRI paradigm control, physiology, and behavior combined with near real-time statistical analysis. NeuroImage 1999; 10:91–106.
3. Cox RW, Jesmanowicz A. Whole brain real-time fMRI. In: Proceedings, ISMRM, 6th Annual Meeting, Sydney, 1998. p 295.
4. Goodyear BG, Gati JS, Menon RS. The functional scout image: Immediate mapping of cortical function at 4 Tesla using receiver phase cycling. Magn Reson Med 1997;38(2):183–186.
5. Gering DT, Weber DM. Intraoperative, real-time functional MRI. J Magn Reson Imaging 1998;8(1):254–257.
6. Goddard NH, Hood G, Cohen JD, Eddy WF, Genovese CR, Noll DC, Nystrom LE. On-line analysis of functional MRI datasets on parallel platforms. J Supercomputing 1997;11:295–318.
7. Cox RW, Jesmanowicz A, Hyde JS. Real-time functional magnetic resonance imaging. Magn Reson Med 1995;33:230–236.
8. Bandettini PA, Jesmanowicz A, Wong EC, Hyde JS. Processing strategies for time-course data sets in functional MRI of the human brain. Mag Reson Med 1993;30:131–176.
9. Watanabe T, Harner AM, Miyauchi S, Sasaki Y, Nielsen M, Palomo D, Mukai I. Task-dependent influences of attention on the activation of human primary visual cortex. Proc Natl Acad Sci USA 1998;95:11489–11492.
10. Kastner S, De Weerd P, Desimone R, Ungerleider LG. Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science 1998;282:108–110.
11. Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ. Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 1997;6:218–229.
12. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes in C, the art of scientific computing, 2nd ed. New York: Cambridge University Press; 1995.
13. Trefethen LN, Bau D. Numerical linear algebra. Philadelphia: SIAM; 1997.
14. Sorenson HW. Kalman filtering: theory and application. New York: IEEE Press; 1985.
15. Sorenson HW, Sacks JE. Recursive fading memory filtering. Information Sci 1971;3:101–119.
16. Ng MK, Plemmons RJ. Fast recursive least squares adaptive filtering by fast Fourier transform-based conjugate gradient iterations. SIAM J Sci Comput 1996;17:920–941.
17. Friston KJ, Jezzard P, Turner R. Analysis of functional MRI time-series. Human Brain Mapping 1994;1:153–171.
18. Rajapakse J, Kruggel F, von Cramon DY. Neuronal and hemodynamic responses from functional MRI time-series: A computational model. In: Kasabov N, Kozma R, Ko K, O'Shea R, Coghill G, Gedeon T, editors. Progress in connectionist-based information systems (ICONIP'97). Singapore: Springer; 1997. p 30–34.
19. Aguirre GK, Zarahn E, D'esposito M. The variability of human, BOLD hemodynamic responses. Neuroimage 1998;8(4):360–369.
20. Cohen MS. Parametric analysis of fMRI data using linear system methods. NeuroImage 1997;6:93–103.
21. Vazquez AL, Noll DC. Nonlinear aspects of the BOLD response in functional MRI. Neuroimage 1998;7:108–118.
22. Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 1996;29:162–173.
23. Ward BD, Garavan H, Ross TJ, Bloom AS, Cox RW, Stein EA. Nonlinear Regression for FMRI time series analysis. In: Proceedings, 4th International Conference on Functional Mapping of the Human Brain, Montreal, 1998. p 767.
24. Shewchuk JR. An introduction to the conjugated gradient method without the agonizing pain. In: FTP: REPORTS.ADM.CS.CMU.EDU (IP: 128.2.222.79) 1994/CMU-CS-94-125.ps
25. Schor S. Diploma thesis, Echtzeitverarbeitung von funktionaler Kernspin-Tomographie, Giessen/Friedberg (Germany), 1997.
26. Schor S, Gembris D, Taylor JG, Peyerl M, Müeller E, Posse S. Functional imaging in real-time (FIRE). In: Proceedings, ISMRM, 6th Annual Meeting, Sydney, 1998. p 1140.
27. Gembris D, Taylor JG, Schor S, Kiselev V, Suter D, Posse S. Methodology of fast correlation analysis for real-time fMRI experiments. In: Proceedings, ISMRM, 6th Annual Meeting, Sydney, 1998. p 1486.
28. Lowe MJ, Russell DP. Treatment of baseline drifts in fMRI time series analysis. J Comp Assisted Tomography 1999;23(3):463–473.
29. Menon RS, Luknowsky DC, Gati JS. Mental chronometry using latency-resolved functional MRI. Proc Natl Acad Sci USA 1998;95(18):10902–10907.
30. Posse S, Schor S, Gembris D, Müller E, Peyerl M, Kroeker R, Grosse-Ruyken ML, Elghahwagi B, Taylor JG. Real time fMRI on a clinical whole body scanner—Single trial detection of sensorimotor stimulation and visual recall activation. In: Proceedings, ISMRM, 6th Annual Meeting, Sydney, 1998. p 162.
31. Schneider F, Weiss U, Salloum JB, Posse S. Real time analysis of amygdala activation. NeuroImage 1999;9(6):S534.
32. Gorse D, Shepherd AJ, Taylor JG. The new ERA in supervised learning. Neural Networks 1997;10(2):343–352.
33. Barhen J, Protopopescu V, Reister D. TRUST: A deterministic algorithm for global optimization. Science 1997;276:1094–1097.