NeuroImage **10,** 530–543 (1999)
Article ID nimg.1999.0490, available online at http://www.idealibrary.com on IDEAL®

Comparison of Filtering Methods for fMRI Datasets

F. Kruggel, D. Y. von Cramon, and X. Descombes

Max-Planck-Institute of Cognitive Neuroscience, Stephanstrasse 1, 04103 Leipzig, Germany

Received August 31, 1998

When studying complex cognitive tasks using functional magnetic resonance imaging (fMRI) one often encounters weak signal responses. These weak responses are corrupted by noise and artifacts of various sources. Preprocessing of the raw data before the application of test statistics helps to extract the signal and can vastly improve signal detection. Artifact sources and algorithms to handle them are discussed. In an empirical approach targeted to yield an optimal recovery of the hemodynamic response, we implemented a test bed for baseline correction and noise-filtering methods. A known signal is modulated onto foreground patches obtained from event-related fMRI experiments. Quantitative performance measures are defined to optimize the characteristics of a given filter and to compare their results. Marked improvements in the sensitivity and selectivity are achieved by optimized filtering. Examples using real data underline the usefulness of this preprocessing sequence. © 1999 Academic Press

Key Words: **fMRI, spatio-temporal filtering, physiological noise, preprocessing.**

INTRODUCTION

Many neuronal brain activations elicit an oxygen consumption and give rise to a hemodynamic response (HR) of the supplying vascular system. This response is measured as the so-called BOLD (blood-oxygen-level-dependent) effect in functional magnetic resonance imaging (fMRI). In T2*-weighted images, this BOLD effect gives rise to an intensity change restricted to a certain brain area and transient in time with respect to the stimulation. However, effects are small and corrupted by noise and artifacts, so that roughly 40–200 repetitions are necessary to detect a statistically significant response. This high repetition count is especially problematic with complex cognitive tasks like memory or language experiments, and one obviously wants to limit the number of repetitions with the help of more sensitive signal detection methods.

In addition to thermal noise, the following artifact sources have been identified as corrupting the BOLD signal and interfering with the analysis of fMRI data: (1) gross body movements during the experimental session, (2) physiological movements (pulsations, swallowing, abdominal movements, breathing), (3) regional sensitivity losses (signal voids) due to susceptibility differences at tissue borders (i.e., the transition bone–brain), (4) ghost images, (5) flow artifacts in the vicinity of large vessels, and (6) long-term instabilities of the scanner baseline. Depending on their time scale, these sources appear as "trends," "fluctuations," or "noise" in the time series; depending on their origin, they may be classified as "physiological" or "system" noise. Of course, other cognitive processes unrelated to the experimental task contribute to the physiological noise.

The best approach for artifact reduction is to work directly on their sources, i.e., improve the hardware, the scanning protocol, and subject fixation. Some major sources are hard to avoid, like heartbeat and breathing (Weisskoff *et al.,* 1993; Jezzard *et al.,* 1993; Hu *et al.,* 1995; Biswal *et al.,* 1996; Noll *et al.,* 1996; Scarth *et al.,* 1996; Le *et al.,* 1996), swallowing (Birn *et al.,* 1998), or acustical scanner noise (Bandettini *et al.,* 1998). At any stage of technical development, some amount of artifacts is found in the recorded data, which may be handled by *post hoc* correction methods. They (1) use prior knowledge about the spectral characteristics of the artifact sources (Weisskoff *et al.,* 1993; Jezzard *et al.,* 1993; Bandettini *et al.,* 1998; Lowe *et al.,* 1998), (2) apply corrections from a biosignal recorded alongside (Hu *et al.,* 1995; Biswal *et al.,* 1994, 1996; Noll *et al.,* 1996; Birn *et al.,* 1998), or (3) make use of multiple image acquisitions (Hu *et al.,* 1994; Le *et al.,* 1996; Buonocore *et al.,* 1997; Sijbers *et al.,* 1998).

Instead of focusing on the best correction method for any given artifact source we may pragmatically ask for the best method to retrieve the BOLD signal from fMRI data. Following a strategy from classical signal analysis, it is advantageous to separate a signal conditioning step (i.e., the deconvolution of artifacts) from the signal detection step (i.e., the inferential statistics). The challenge of fMRI preprocessing is (1) to separate the task-related BOLD signal from ongoing artifacts and noise unrelated to the task (and thereby increasing the sensitivity), (2) to recover the BOLD signal shape, and

1053-8119/99 $30.00

(3) to retain the high spatial resolution of fMRI. The objective of this paper is to compare several filtering methods for their use in fMRI preprocessing using a filter testbed. Based on quantitative comparisons, we derive conclusions about their usefulness in a baseline correction and signal restoration step.

TEST BED DESCRIPTION

To derive conclusions based on quantitative evaluations, we constructed a test bed, in which we modulated a known signal with a known spatial distribution onto patches obtained from fMRI experiments presumed to contain only background noise. We compared spatiotemporal properties of the known signal with the filter output and characterized the filter performance by six different test measures.

Our primary interest is in processing fMRI data using event-related experimental designs. Throughout this paper, we assume that the HR due to an experimental stimulus is monophasic, i.e., the fundamental frequency of the BOLD signal corresponds to the stimulation frequency ν. We found this assumption to be valid in all experimental designs studied so far.

Most fMRI protocols available today allow the acquisition of a set of image slices at a given time step, and typically, there is a spacing (of 2 mm) between slices. We consider slices to be independent and treat two-dimensional spatial models only. All algorithms discussed here are easily extensible to three spatial dimensions. In our models the third dimension corresponds to time.

Test Patches

To cover the range of typical event-related paradigms in cognitive fMRI experiments, we have selected three different studies: (1) an experiment of language comprehension (TR 2 s, 12 time steps per trial, 76 trials), (2) a working memory experiment (TR 1 s, 18 time steps per trial, 104 trials), (3) an experiment in lexical priming (TR 1 s, 8 time steps per trial, 320 trials). All experiments were run using an EPI protocol on a Bruker 3T Medspec 300 system (64×64 voxel, resolution $3.8 \times 3.8 \times 5$ mm, 2-mm gap, axial slices parallel to AC–PC). All further slice positions are given relative to the AC–PC plane.

From each of these studies, four datasets were randomly chosen. Two patches of 10×10 voxels were selected from each dataset at a brain region where no activation was found using a standard statistical procedure (see below). In the case of aurally presented stimulation material (experiments 1 and 3), these patches were taken from the right occipital lobe (at +0 mm) and the right frontal lobe (+24 mm). In experiment 2, visual stimulation was used, so we selected patches from the right frontal lobe (+0 mm) and a right centroparietal region (+42 mm). All patches were free of background voxels, but contained a random proportion of white matter, gray matter, and CSF voxels. A total of 24 patches were included in this study.

For detection of functional activation, we took a stimulus function derived from the experiment, convolved it with a Gaussian hemodynamic modulation function using a lag of 4.8 s and a dispersion of 3.6 s (Clark *et al.,* 1998), computed a Pearson correlation with the convolved waveform, transformed the correlation coefficient into a z score, and determined the significance of activated clusters by their spatial extent for $z > 3$ (Friston *et al.,* 1994).

Test Signals

Because the spectral properties of the HR are unknown beforehand, we included three different waveforms as BOLD signal models in our test suite: (1) a sine wave; (2) a prototypical HR, which was obtained by averaging the responses of a highly activated sample region in space and time; and (3) a square wave (see Fig. 1). The properties of a real BOLD signal are expected to lie within these extremes. The period length was matched against the trial length of the corresponding experiment.

Spatial Pattern

We defined two spatial patterns for the test signal, a single 4-voxel activation and four single voxel activations (see Fig. 1). We used sharp transitions between activated and nonactivated regions in order to test the spatial properties of a filter. For a specific test, one of the three waveforms was modulated onto this patch using one of the two spatial patterns, multiplied by an amplitude factor to achieve a certain signal-to-noise ratio on input.

Test Measures

The performance of the filter methods was determined by a set of test measures which describe the spatial and temporal properties, the sensitivity and selectivity of a filter. First, we need to give a few basic definitions:

The signal energy E_{sig} is computed from a signal $y(t)$ sampled regularly at N time steps via its periodogram $P(\cdot)$ (see Press *et al.,* 1992, pp. 549–553) by summing over the multiples of the base frequency f:

$$E_{\text{sig}} = \sum_{k=1}^{N/(2f)} P(kf). \tag{1}$$

Because additive noise (and perhaps some subtle undetected activation) in the patch contributes to E_{sig}, we define a "true" signal energy as the difference between the signal energy and the spectral energy E_0 at the

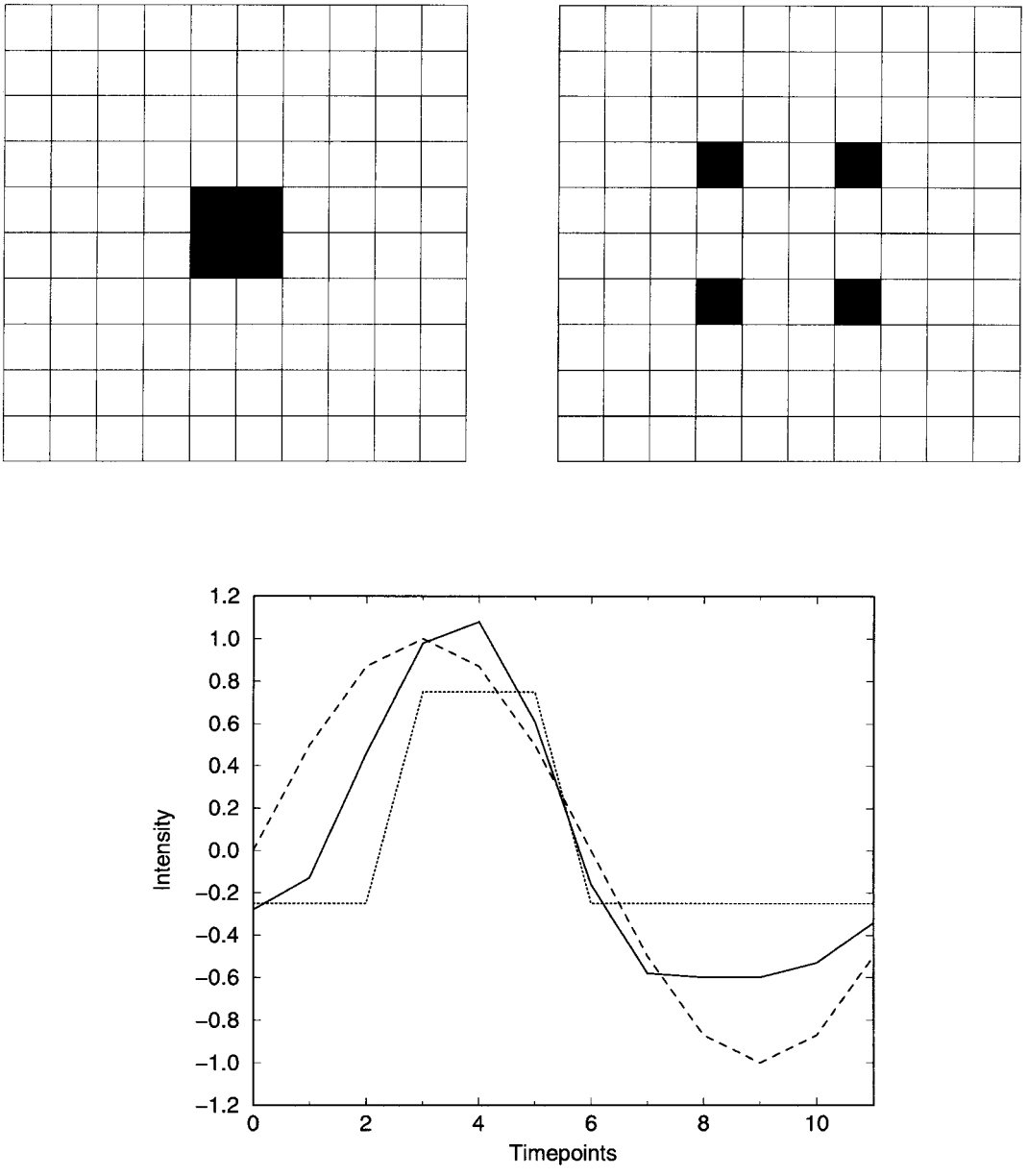


FIG. 1. Top: Scheme of the two spatial patterns of the test signal. A test signal is modulated on black pixels, white pixels are untouched. Bottom: Test waveforms applied for patches from experiment 1. Sine wave (dashed), square wave (dotted), and prototypical HR (solid).

same frequencies if no modulated signal is present: $E_{\text{true}} = E_{\text{sig}} - E_0$. The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = E_{\text{true}} \Big/ \sum_{k=0}^{N/2} P(k). \tag{2}$$

The correlation C_{ab} of signals $y_a(t)$ and $y_b(t)$ is given by the product of their normalized periodograms:

$$C_{ab} = \sum_{k=0}^{N/2} P_a(k) P_b(k). \tag{3}$$

Now we define the following test measures:

- The *recovery* of a preprocessing method is defined as the correlation C_{to} between the test signal and the output signal. The recovery ranges between 0 and 1, where 1 corresponds to a perfect restoration of the test signal. This measure defines a temporal (shape) restoration property of a preprocessing method.
- By *sensitivity* we denote the least SNR for which the signal can be detected by a standard statistical procedure (see Test Patches). We compute a Pearson correlation of the output signal with the test function, convert the correlation coefficient into a z score, and average across the foreground voxels. Sensitivity val-

ues are given for $\bar{z} = 3.0$. Thus, lower SNRs indicate a better filter sensitivity.

- *Selectivity* denotes how well a filter selects the test signal: The test signal was scrambled randomly in time and modulated onto the patch, and its energy E_{ran} after preprocessing was determined. The selectivity is then computed as E_{sig}/E_{ran}.
- The ratio SNR_{out}/SNR_{in} should ideally be independent of the input signal level. When comparing filter characteristics, the absolute value of this ratio is less relevant than the range of signal levels for which this ratio is maintained constant. We define *linearity* as the range of signal levels for which the change of the ratio is less than 10%. For quantitative analysis of fMRI data, filter linearity is an important feature.
- By *spatial blurring B* we define a measure of how much a spatial filter blurs the spatial modulation pattern: $B = \overline{E}_0/\overline{E}_1$, where $\overline{E}_0$ denotes the mean true energy of the background ($\overline{E}_1$, foreground) voxels of the test pattern.
- Because we modulate the same signal energy on all foreground voxels, an ideal filter should yield the same relative level on output. The *spatial smoothness S* is defined as the relative energy spread of the output signal in the set of foreground voxels: $S = (E_{max} - E_{min})/\overline{E}$, where E_{min} denotes the minimal and E_{max} the maximal true signal energy and $\overline{E}$ their mean. This measure defines a spatial restoration property of a preprocessing method: low values of S indicate a higher regularization and thus a better restoration.

BASELINE CORRECTION

In fMRI datasets, intensity fluctuations modulate the time series of any foreground voxel (see Fig. 2, top). These fluctuations are slow with respect to the length of an experimental trial, i.e., their frequency is lower than the stimulation frequency. They depend on the voxel location (i.e., on the tissue properties), on the scanner properties (i.e., linearity, long-term stability), and perhaps on physiological effects. These fluctuations often make up to 10% of the total signal intensity and thus can easily hide functional activations. The (time-independent) intensity of a foreground voxel, together with the fluctuations, is referred to as baseline. Trends and fluctuations violate the assumption of temporal stationarity, which is a requirement for most test statistics.

To correct for baseline fluctuations, the easiest way is to estimate the baseline and subtract it from the input signal. Here, one has to be cautious not to model signal changes attributable to functional activation. In the temporal domain, any suitable baseline estimator has the spectral properties of a low-pass filter with a cutoff frequency below the stimulation frequency.

There is a broad range of algorithms available from classical time series analysis for baseline estimation. We selected four typical examples: (1) moving average (MA) filters, (2) low-pass (LP) finite impulse response filters, (3) autoregressive filters (AR models), and (4) Kalman filters. Subtraction of the filter output from the input signal effectively results in high-pass filtering. The anatomical information (i.e., the time-independent part) is separated from the functional activation. We tested the performance of the four baseline filters for all 24 patches from the three experiments, using test signals at SNR = 0.2. To conserve space, figures shown here correspond to test runs using a prototypical hemodynamic response, averaged across results from all 24 test patches, and include standard error bars.

Moving Average Filter

The MA filter estimates the state of a voxel at time t_0 from the mean in a window of $2N + 1$ time points around t_0:

$$y_s^f(t_0) = \sum_{r=-N}^{N} y_s(t_0 + r)/(2N + 1), \qquad -N \le r \le N. \quad (4)$$

The MA filter is characterized by its window size, or filter length. Example results from the test bed for the prototypical HR using the single peak pattern are compiled in Fig. 3.

The MA filter showed a performance optimum for all test measures at a filter length which corresponds to approximately 1.4 times the length of a single trial, i.e., for a trial length of 12 time steps, averaging over 17 time steps achieved optimal results for baseline estimation. Using fewer coefficients spoiled the functional activation while longer filters did not follow baseline fluctuations close enough.

Finite Impulse Response Low-Pass Filter

The LP filter is described by (Rorabaugh, 1993)

$$y_s^f(t_0) = \sum_{r=-N}^{N} \phi_r w_r y_s(t_0 + r), \qquad -N \le r \le N, \quad (5)$$

where ϕ_r denotes the $2N + 1$ low-pass filter coefficients for the cutoff frequency λ,

$$\phi_r = \begin{cases} \lambda/\pi & \text{if } r = 0, \\ \sin(r\lambda)/\pi & \text{if } 0 < |r| \le N, \end{cases} \quad (6)$$

and w_r the $2N + 1$ coefficients of a Hamming window,

$$w_r = 0.54 + 0.46 \cos\left(\frac{\pi r}{N}\right), \qquad -N \le r \le N. \quad (7)$$

The LP filter is characterized by its cutoff frequency λ, while $N = 25$. Example results, again using the single peak pattern, are compiled in Fig. 4.

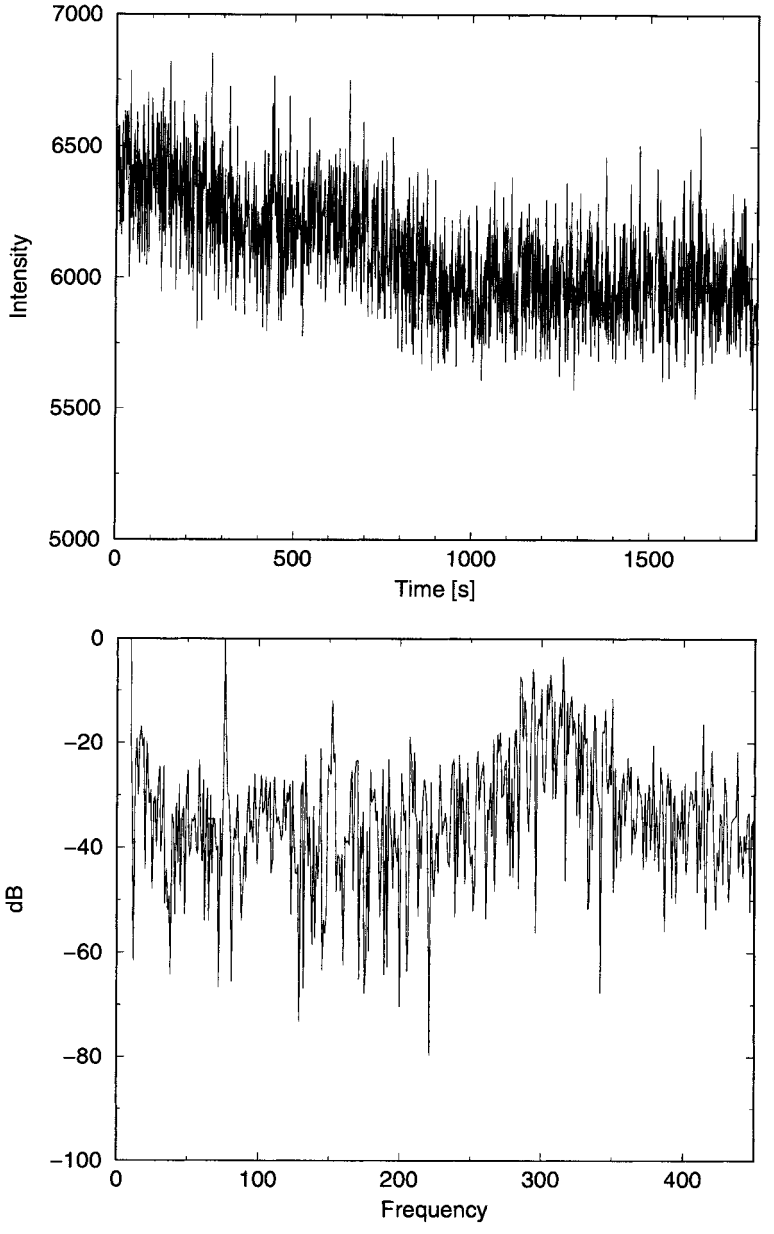


FIG. 2. Native time series (top) and corresponding power spectrum (bottom) of a sample activated voxel. In the time series, baseline fluctuations and a negative drift are visible. The power spectrum exhibits a prominent peak at 76, which corresponds to the cognitive stimulation frequency. The broad peak at higher frequencies is attributed to breathing.

The LP filter exhibited an almost constant performance for cutoff frequencies, which were equal or lower than the stimulation frequency ν.

Autoregressive Filter

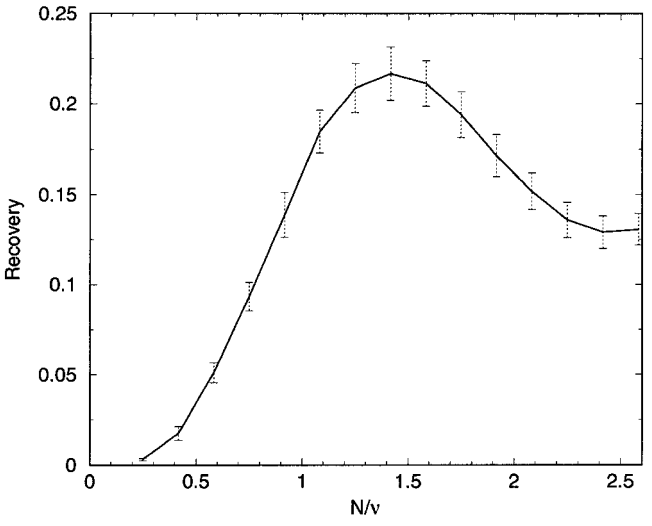
For the AR filter, we consider the baseline to be a realization of the process

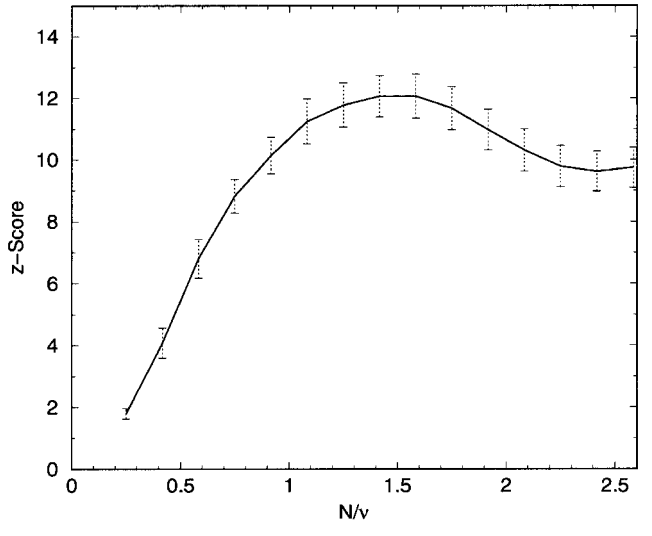
$$y_s^f(t) = \sum_{r=1}^{p} \alpha_r y_s(t - r) + \epsilon(t), \tag{8}$$

where α_r is the p parameter of the AR(p) process (West *et al.,* 1997), and $\epsilon(t) \sim N(0, \sigma^2)$. In the tests, only for a rather high number of coefficients ($p \geq 15$), a performance comparable to the two previous filters was obtained. A comparatively low recovery for the signal shape was found, and the optimum number of parameters depended on the noise characteristics of the patch under study.

Kalman Filter

A stateless Kalman filter (Grewal *et al.,* 1993; Petersen *et al.,* 1998) may also be applied for baseline





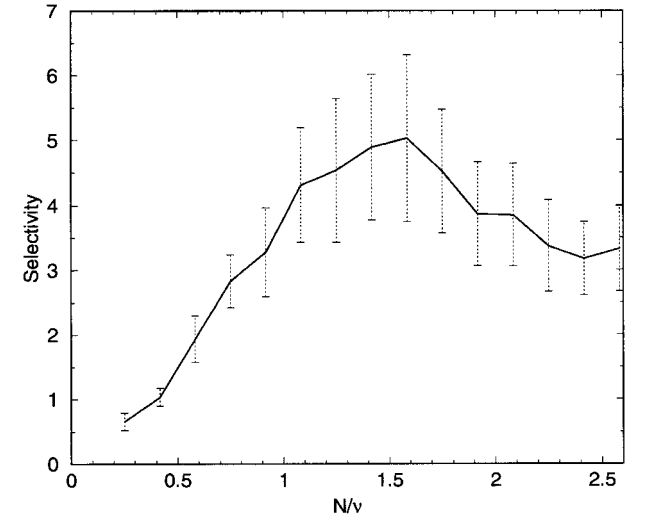
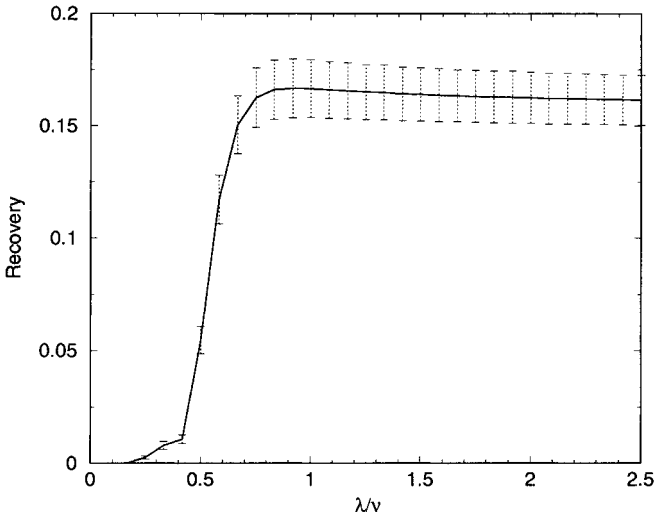
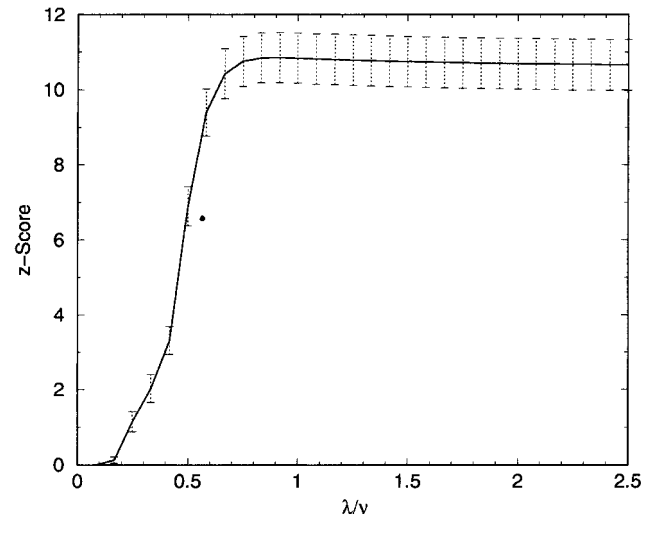


FIG. 3. Recovery (top), z score (middle), and selectivity (bottom) of a MA filter for baseline correction. The ratio of the filter length N to the period length ν is given on the x axis. A clear optimum is seen at a ratio of 1.4.





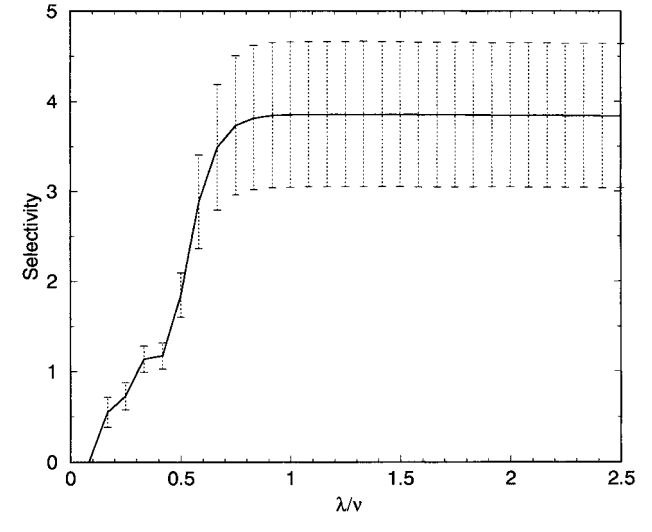


FIG. 4. Recovery (top), z score (middle), and selectivity (bottom) of a LP filter for baseline correction. The ratio of the cutoff frequency λ to the period length ν is given on the x axis.

estimation,

$$y_s^f(t) = y_s(t-1) + (p/(p+r))(y_s(t) - y_s(t-1)), \quad (9)$$

$$p = (1 - p/(p+r))p + q, \quad (10)$$

where p is the estimated covariance, q is the process noise, and r is the measurement covariance. The properties of such a filter depend on the ratio r/q. We set $r = 0.01$ empirically and set $q = r/10^f$, where f is the filter parameter.

The Kalman filter showed a reasonable performance for a filter parameter $f \geq 1$, which corresponds to settings of the measurement covariance $r = 0.01$ and the process noise $q = 1e - 3$.

Filter Comparison

From the results of the test bed, we determined optimal parameters for each filter. We modulated waveforms with an input SNR of 0.01–0.20 onto the patches and computed the SNR on output (see Fig. 5) using the best filter parameters. Except for the AR filter, a linear behavior was found. The best sensitivity was determined for the MA filter, closely followed by the LP filter. In Table 1, numerical values for the optimal filter parameters are compiled. Because these filters operate in the temporal domain, the parameters spatial blurring and smoothness are not meaningful here.

Discussion

In addition to removing the T2-weighted image, LP, MA, and Kalman filters achieved a marked increase in the sensitivity and selectivity of the signal detection as well as in the recovery of the signal shape. Although the peak performance of a MA filter was better than the Kalman and LP filters with respect to all parameters, the latter reveals a much higher independence of the results in relation to the filter parameter. Thus, the LP filter does not need to be fine-tuned with respect to the stimulation frequency. Because the AR filter is computationally intensive and exhibits a nonlinear behavior, baseline estimation by an AR model is not considered useful.

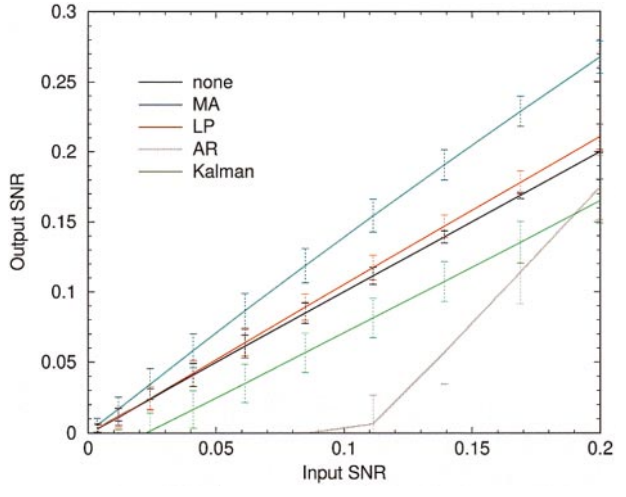


FIG. 5. Linearity of baseline correction methods. The input SNR is plotted against the output SNR. All baseline filters were supplied with the best parameter settings. In order of decreasing sensitivity, MA > LP > Kalman > AR.

TABLE 1

Performance of Filters for Baseline Correction

Filter	Parameter	Recovery	Sensitivity	z score	Selectivity
None	—	0.000	0.0720	8.06	2.84
MA	$N/\nu = 1.4$	0.216	0.0427	12.0	4.89
LP	$\lambda/\nu = 1.5$	0.164	0.0471	10.7	3.85
AR	$p = 20$	0.137	0.1294	10.2	8.09
Kalman	$f = 2$	0.115	0.0782	7.28	2.88

Note. Shown are results for optimal filter parameters.

We restricted to the baseline estimation in the time series only: a spatiotemporal filter needs to track anatomical edges (i.e., the transition CSF–cortex) and thus requires edge-preserving properties in the spatial domain. We tested spatial extensions of a temporal moving average and finite impulse response filter and did not find any advantage in using spatial information here.

Estimation of the baseline by a matched LP filter appears to be sufficient for the correction of baseline fluctuations. In all experiments studied so far, a cutoff frequency of 1.0–1.5 times the trial length (in time steps) appears to be sufficient. By baseline correction we get rid of anatomical edges in the data, which is a prerequisite for efficient spatiotemporal noise filtering.

SIGNAL RESTORATION

After baseline correction, the remaining data are composed of (1) functional activation, (2) physiological noise, and (3) system noise. The aim is to restore the shape of the functional activation in space and time, thus separating it from signal variations uncorrelated with the experimental task. In the temporal domain, this corresponds to filtering higher frequency portions, like the broad peak in the example spectrum of Fig. 2 (bottom), which corresponds to artifacts introduced by breathing.

Because anatomical edges are now absent, we may take advantage of any spatial coherence to restore the functional signal. The aim of restoration is to add some constraints of homogeneity and smoothness on the solution. These constraints may be applied either in the temporal or in the spatial domain, separately, or in both

domains. In the temporal domain, a smoothness constraint is justified by an upper found for the slope of the BOLD signal. The slope of the HR is found on the order of 4–6 s and thought to be limited by vascular response mechanisms. In the spatial domain, prior knowledge about the size of the activated structures may be used as spatial smoothness constraints: for example, the BOLD signals from cortical regions appear as clusters of 3–12 mm width (corresponding to 1–4 voxels).

However, results of the signal restoration step are based on a tradeoff between conflicting aims: to achieve smoothness in the spatial domain without spoiling the activation and to limit the bandwidth in the temporal domain in order to filter out quasiperiodic artifacts without altering the (presumed) HR shape.

We studied five algorithms for signal restoration. The first three operate in the temporal dimension, the fourth operates in the spatial domain, and the last operates in the spatiotemporal domain. We tested the performance of the five restoration algorithms for all 24 patches from the three experiments, using test signals at SNR = 0.2. Prior to signal restoration, all patches were baseline corrected using a LP filter ($\lambda/\nu = 1.5$). To conserve space, figures shown here correspond to test runs using a prototypical hemodynamic response, averaged across results from all 24 test patches, and include standard error bars.

Finite Impulse Response Low-Pass Filter

The LP filter in the temporal domain was already described in the previous section [see Eq. (5)]. For signal restoration, we are interested in the low-frequency band, i.e., we use the filter output directly. Example results from the test bed are shown in Fig. 6.

The shape recovery shows an optimum at $\lambda/\nu = 0.38$, which corresponds to a passband including the fundamental stimulation frequency and its first harmonic. For this setting, plots of the z score and the selectivity are also optimal. Note that this filter setting depends on the stimulation frequency only.

Temporal Gaussian Filter

Gaussian filtering in the temporal domain (GT) is achieved by a convolution of the time series with a one-dimensional Gaussian function:

$$y_s^f(t_0) = \sum_{r=-N}^{N} g_r y_s(t_0 + r), \qquad \text{with } g_r = \frac{1}{\sqrt{2\pi}\sigma} \exp(-r^2/2\sigma^2), \quad -N \le r \le N. \tag{11}$$

The GT filter is characterized by the standard deviation σ of the Gaussian function, with filter dimensions of

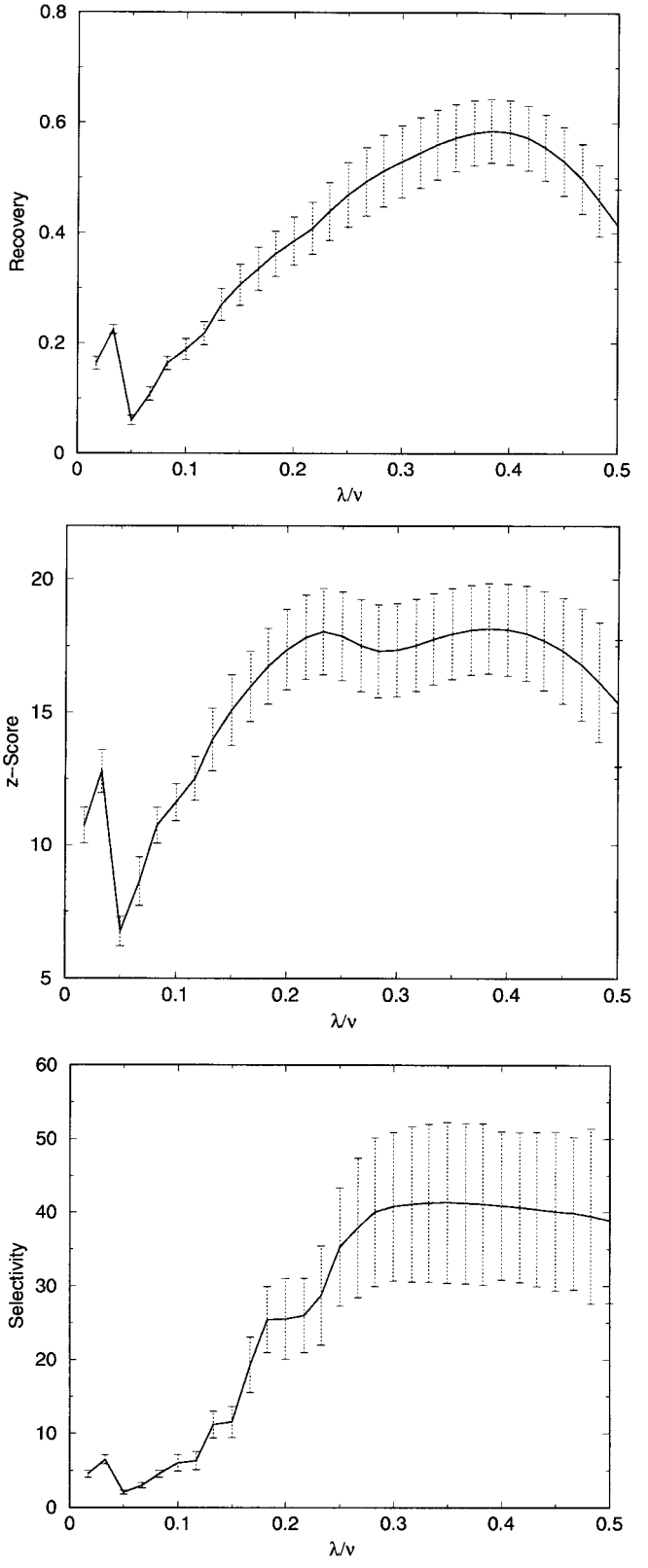


FIG. 6. Recovery (top), z score (middle), and selectivity (bottom) of a LP filter for signal restoration. The ratio of the filter length λ by the period length ν is given on the x axis. An optimum is seen at a ratio of 0.38.

typically $N \geq 3\sigma$. Example results for comparison with the LP filter are given in Fig. 7.

We found a rather broad peak at $\sigma = 1.8$ in the signal recovery, which corresponds to the temporal smoothness of the HR (at a TR = 2 s here). The maximum z score was located at $\sigma = 1.4$, while the selectivity had a broad maximum starting at $\sigma = 1.6$. Thus, a setting of $\sigma \geq 1.6$ emerged as a good compromise here. Note that this setting is independent of the stimulation frequency; however, it needs to be matched against the slope of the HR, which depends on the stimulus characteristics, the brain region, and the individual physiological factors. In the three different experiments studied here, an optimal setting range was $\sigma = 1.4$–2.0.

Autoregressive Filter

An AR process may be used to model the functional activation in the temporal domain directly. The computational outline was already given in Eq. (8). A clear peak was found for an AR(2) model. However, in comparison with the two previous filters, performance was low for the recovery and selectivity. Again, a nonlinear behavior was found (see below). We also tested ARMA models, but did not obtain any advantage over AR filters.

Spatial Gaussian Filter

Gaussian filtering in the spatial domain (GS) is a popular choice for noise filtering, especially in fMRI preprocessing (Lowe and Sorenson, 1997). GS filtering is achieved by successive convolutions in both spatial dimensions. The filter is characterized by the parameter σ, which is converted into the other well-known characterization, the full width at half-maximum (FWHM) by multiplying by 2.34. The spatial properties of this filter are shown in Fig. 9.

In terms of the performance measures, all parameters decreased dramatically with increasing σ to lower values than the unprocessed case, which is explained by the spatial blurring induced by this filter. Only the spatial smoothness revealed a shallow optimum at $\sigma = 1.2$ (FWHM = 2.8 voxels). At this setting, there was considerable blur and loss of sensitivity and selectivity. The recovery decreased by a third of the unprocessed case, and was 10 times lower in comparison with other methods (see Table 2). We consider GS filtering not useful for the restoration of small signals in event-related fMRI data.

Signal Restoration by a Markov Random Field Model

We included a filter based on a Markov random field (MRF) to introduce a spatiotemporal model for signal restoration. With respect to a comprehensive description in this journal (Descombes *et al.,* 1998), we deliberately keep the discussion short here.

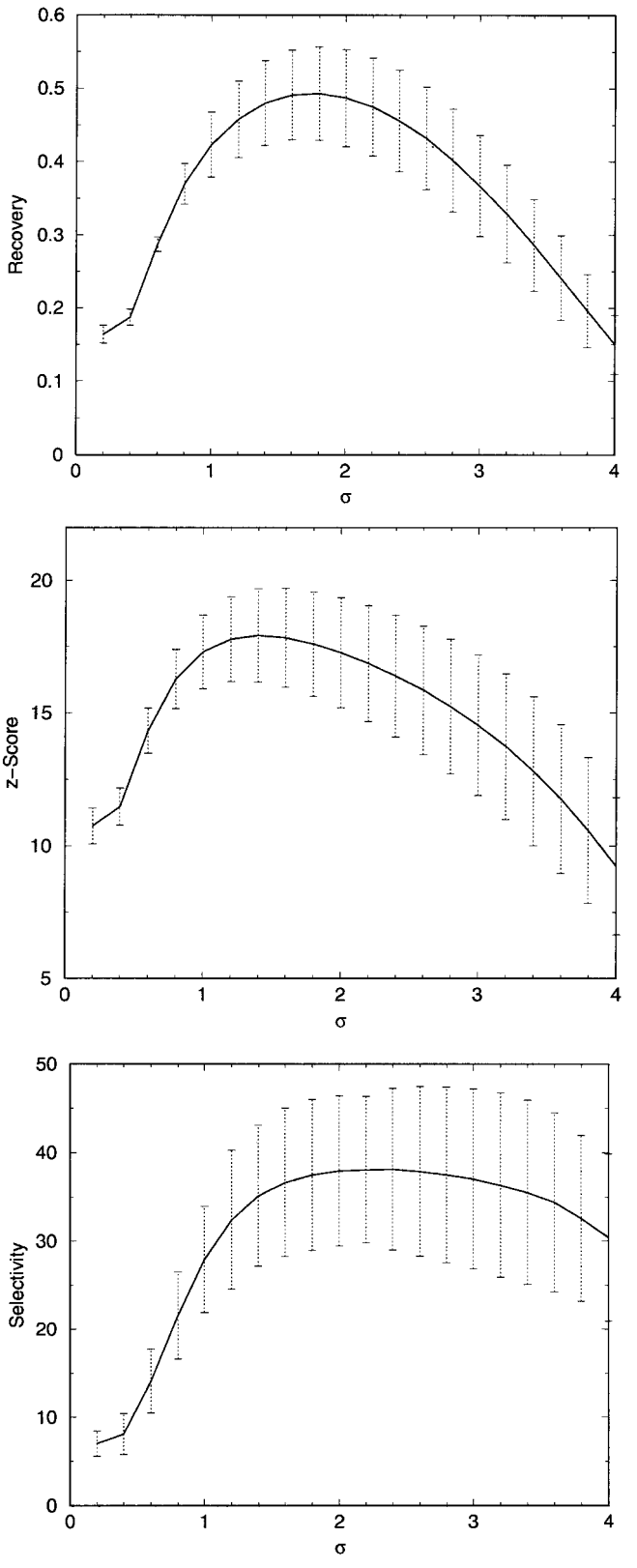


FIG. 7. Recovery (top), z score (middle), and selectivity (bottom) of a Gaussian filter in the temporal domain for signal restoration. The filter parameter σ is given on the x axis. For this experiment, TR = 2.0 s.

TABLE 2

Performance of Filters for Signal Restoration

Filter	Parameter	Recovery	Sensitivity	z score	Selectivity
None	None	0.164	0.0471	10.7	6.98
LP	$\lambda/\nu = 0.38$	0.584	0.0189	18.1	40.0
GS	$\sigma = 1.2$	0.054	0.0929	6.01	3.10
GT	$\sigma = 1.6$	0.491	0.0219	17.8	36.6
AR	$p = 2$	0.151	0.0471	9.65	7.47
MRF	$\beta = 1.0, \delta = 1600$	0.616	0.0151	20.7	18.7

Note. Shown are results for the optimal filter parameters.

We consider that data X consist of the underlying signal Y corrupted by additive noise η. Reformulating the problem of reconstructing Y given X in a Bayesian framework, we have to maximize the *a posteriori* probability $P(Y|X) \propto P(Y)P(X|Y)$, where $P(Y)$ refers to the prior model and $P(X|Y)$ to the likelihood of model and data. A Gibbs field formulation of the prior probability is given by

$$P(Y) = \frac{1}{Z} \exp\left[-\sum_{c \in C} V_c(y_s(t))\right], \quad (s,t) \in c), \qquad (12)$$

where C is a finite subset of voxels (a clique), V_c is the potential associated with the clique c, and Z is a normalization constant. We consider a 3D spatiotemporal MRF model, in which two dimensions correspond to the spatial dimensions of a fMRI slice and the third dimension refers to the temporal domain. For each voxel site s at time point t, a clique is defined by its four closest spatial neighbors, the two next temporal neighbors, and the two temporal neighbors one stimulation period apart. The interaction between two voxel sites (y_i, y_j) is modeled by the following potential function:

$$V_c(y_i, y_j) = -\beta/(1 + (|y_i - y_j|/\delta)^2). \qquad (13)$$

The β parameter may roughly be understood as the interaction strength between two sites. The δ parameter defines the intensity difference of a transition between activated and nonactivated states. We use the same parameter values in the spatial and the temporal domains. A similar interaction is introduced for the likelihood term,

$$P(X|Y) = \frac{1}{Z} \exp\left[-V_l(y_s(t))\right], \qquad \text{using} \qquad (14)$$

$$V_l(y_s(t)) = -\beta/(1 + (|y_s(t) - x_s(t)|/\delta)^2), \qquad (15)$$

which attaches the restored signal Y to the measured data X. The global energy $U_X(Y)$ for a configuration is

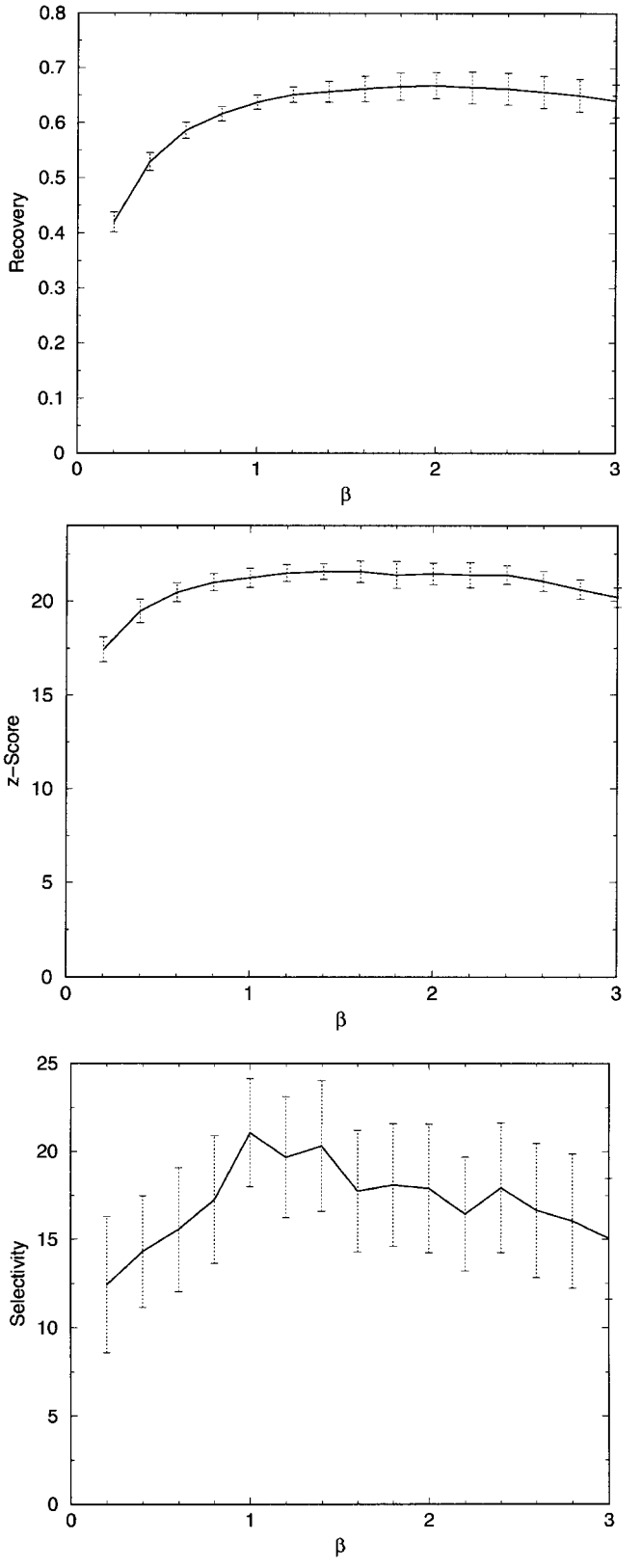


FIG. 8. Recovery (top), z score (middle), and selectivity (bottom) of a MRF model for signal restoration. The filter parameter β (at $\delta = 1600$) is given on the x axis. Note that recovery and z score values exceed those of the previous algorithms.

defined by

$$U_X(Y) = \sum_{s,t} V_l(y_s(t)) + \sum_{c \in C} \sum_{s,t} V_c(c). \qquad (16)$$

To find the configuration which minimizes the global energy, we applied a simulated annealing scheme (Berthod *et al.,* 1996). The spatial properties of the MRF filter are compared with the GS filter in Fig. 9.

Stable performance plots were found for $\delta \geq 800$, which roughly corresponds to 1.5 times the noise level determined in these patches. Recovery, sensitivity, and selectivity benefit from increasing values for the interaction strength β (see Fig. 8). The edge-preserving properties of this filter are documented in spatial blurring values, which were a magnitude better compared with a GS filter (see Fig. 9). From a series of tests, we selected $\delta = 1600$ and $\beta = 1.0$. The stochastic nature of the optimization algorithms resulted in a certain run dependency, which was negligible for this parameter setting.

Filter Comparison

From the results of the test bed, we determined optimal parameters for each filter. We modulated waveforms with an input SNR of 0.01–0.20 onto the patches and computed the SNR on output (see Fig. 10) using the best filter parameters. Except for the AR filter, a linear behavior was found. The best sensitivity was determined for the MRF filter and comparable values for the LP, GS, and AR filters. In Table 2, numerical values for the best filter parameter settings were compiled.

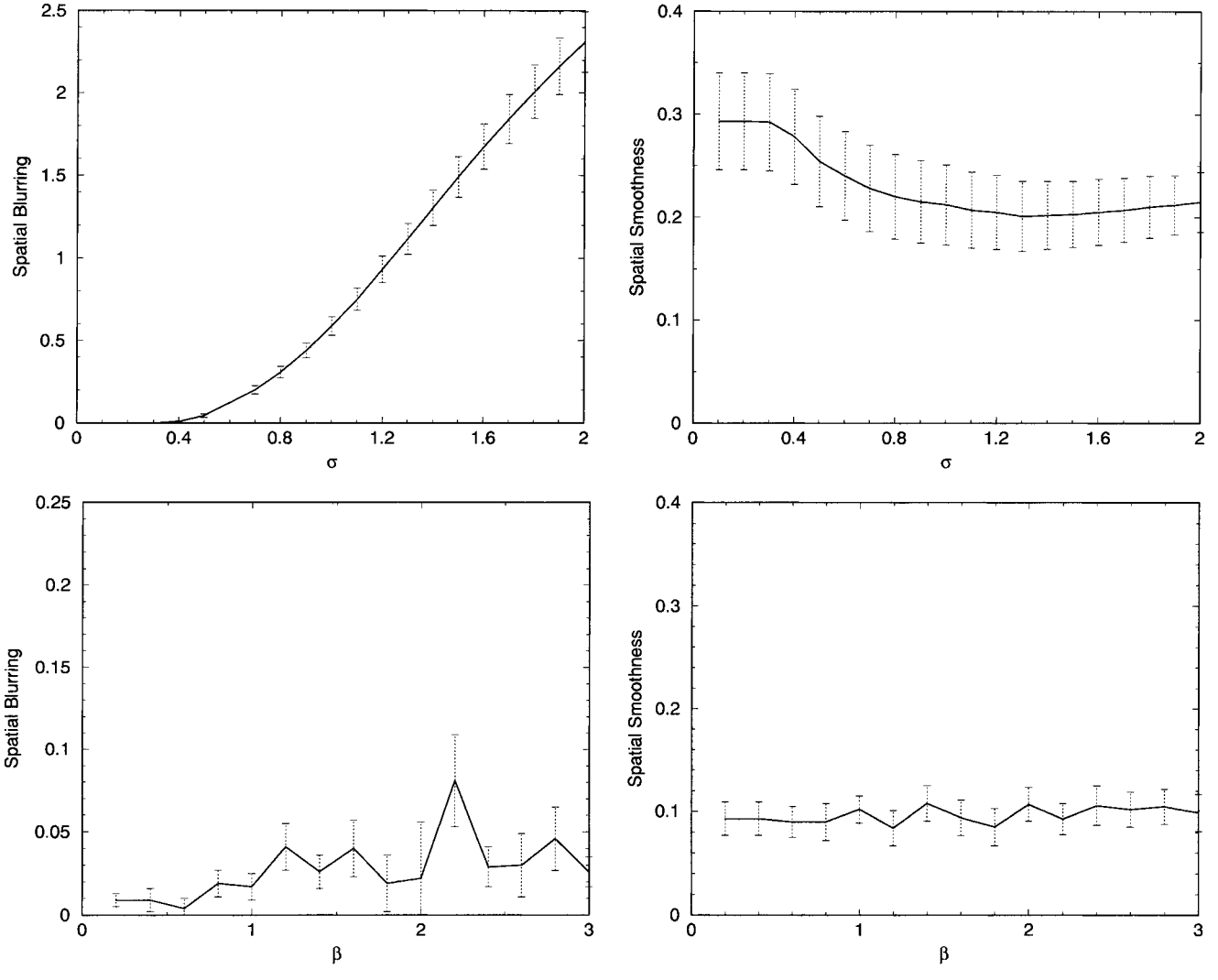


FIG. 9. Comparison of the spatial properties of the GS (top) and MRF filter (bottom). For the GS filter, an optimum is found with $\sigma = 1.2$; however, there is considerable blur and sensitivity loss at this setting. For the MRF filter, the interaction coefficient β is given on the x axis at $\delta = 1600$. The edge-preserving properties of this filter are illustrated by low blurring values and a good spatial smoothness. Note that the scale for the spatial blurring plot is 10 times less than that for the MRF model.

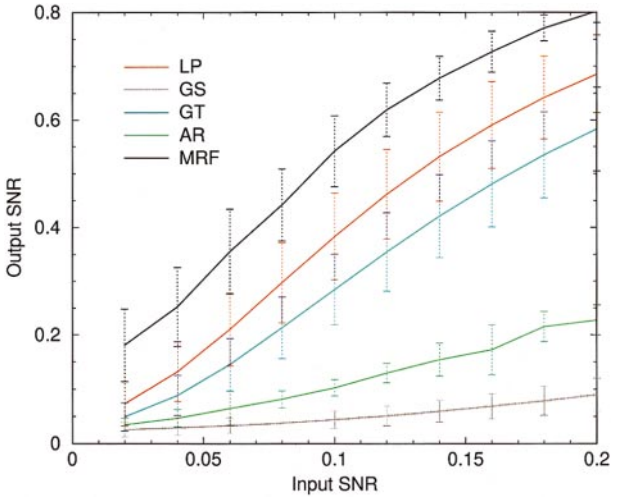


FIG. 10. Linearity of signal restoration methods. The input SNR is plotted against the output SNR. All filters use the best parameters for this problem (see Table 1). In order of decreasing sensitivity, MRF > LP > GT > AR > GS.

Discussion

We summarize the results of the signal restoration tests as follows:

- All filters except GS and AR offered a dramatic increase in all test measures compared with the native case.
- The best figures were obtained with the MRF, LP, and GT filters. Shape recovery and sensitivity were best for the MRF filter, while LP and GT filters provided a better linearity and selectivity.
- In the light of the challenges for preprocessing, best results were obtained with the MRF filter. If the computation time is scarce, the use of a matched LP or GT filter is suggested.

A REALISTIC EXAMPLE

For a final demonstration, we include an evaluation of experiment 1 using baseline correction (LP filter, $\lambda/\nu = 1.5$) and no signal restoration (Fig. 11a) or signal

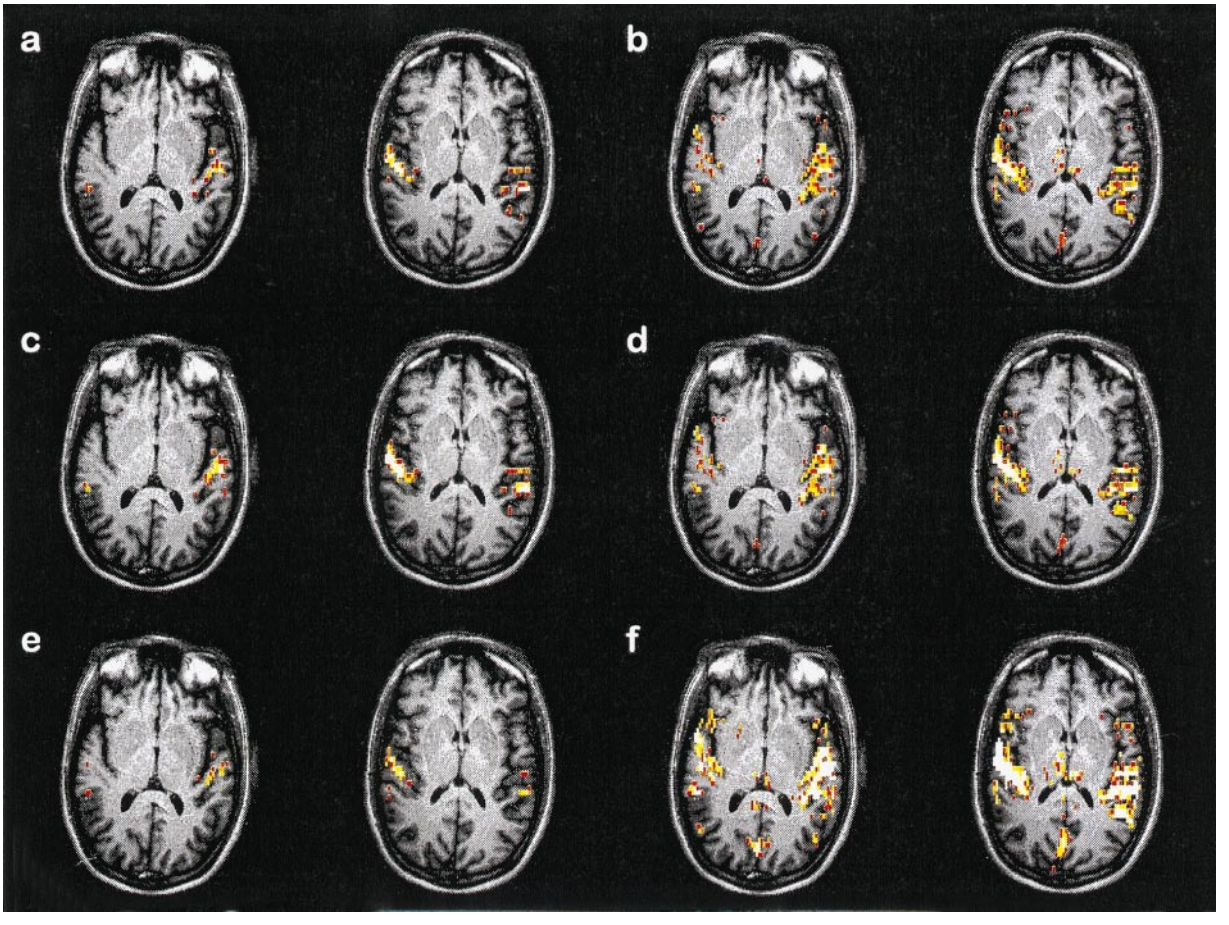


FIG. 11. Comparative evaluation of the different signal restoration methods: (a) native, (b) LP filter, (c) GS filter, (d) GT filter, (e) AR filter, and (f) MRF filter.

restoration using an LP filter (Fig. 11b, $\lambda/\nu = 0.38$), GS filter (Fig. 11c, $\sigma = 1.2$), GT filter (Fig. 11d, $\sigma = 1.6$), AR filter (Fig. 11e, $p = 2$), or MRF filter (Fig. 11f, $\beta = 1.0$, $\delta = 1600$). From the preprocessed datasets, we computed a z score by a standard statistical procedure (see Test Patches). Shown are color-coded overlays with a positive z score range from 12 to 24. A marked increase in sensitivity is found with the LP, GT, and MRF filters, especially for the much better delineated thalamic activation.

GENERAL DISCUSSION

In this paper, we discuss artifact sources corrupting the BOLD signal in fMRI data and how they can be deconvolved using digital filters. With a focus on data obtained in cognitive experiments using event-related designs, special attention has been paid to the recovery of BOLD signals, which are small in amplitude and spatial extent.

It is not unreasonable to assume that results obtained in this study are transferrable to other imaging protocols and to scanners operating at field strengths other than 3T: artifact sources remain the same, only the amount of their contribution may vary. Our test bed allows a transparent and quantitative comparison across filter algorithms and imaging protocols, and we make the test bed available to ease such comparisons.

Filtering techniques discussed here are applicable to fMRI data obtained under blocked designs, but are less effective. With increasing length of an experimental block, the filter passband needs to be widened, and it is harder to discriminate between baseline and BOLD signal fluctuations. Extending the recommendations for event-related experiments to designs with block lengths up to 60 s was found to be possible. When very short intertrial intervals (i.e., 4–8 s) are used, artifacts due to breathing become a major problem, and the filtering methods discussed here do not perform sufficiently well. In these cases, correction approaches using simultaneously recorded biosignals (Biswal *et al.,* 1996; Birn *et al.,* 1998) are expected to yield better results.

With cognitive fMRI experiments, there is an increasing interest in the description of the temporal shape and spatial extent of BOLD responses with respect to modifications of the stimulation conditions. It is still under investigation how much may be inferred about the underlying neuronal activation from shape characteristics of the HR. Throughout this paper, we made the simplifying assumption that the paradigm frequency ν corresponds to the fundamental frequency of the BOLD signal. In experiments using a single stimulus per trial, we did not find any convincing example of a multiphasic response. Moreover, the spectral properties of the BOLD signal show that the first harmonic (2ν) contributes only about 10% to the spectral power, while higher harmonics are indistinguishable from noise. This suggests that the spectral passband should be restricted to a rather narrow range. When multiple responses per trial are expected, one may prefer to increase the cutoff frequency in the restoration step.

By filtering, we enhance autocorrelations in the data. As a consequence we have to correct for the effective degrees of freedom in the statistics (Friston *et al.,* 1995). As discussed by Purdon and Weisskoff (1998), it is preferable to use an empirical noise model instead of deriving it from the hemodynamic response function (Friston *et al.,* 1995). fMRI data were found to be AR(1) correlated in space and time (Bullmore *et al.,* 1996; Benali *et al.,* 1997; Kruggel *et al.,* 1999). Thus, autocorrelations may be taken into account via a spatiotemporal covariance matrix in the statistics.

With a reliable baseline correction scheme, it is possible to compare data from different experimental blocks. Matching filter characteristics to the properties of the BOLD signal in a preprocessing step increases the sensitivity and selectivity of BOLD signal detection and improves the recovery of the BOLD signal shape, which is important to detect small activations found in cognitive experiments. Good signal restoration techniques are a prerequisite for modeling hemodynamic responses with parameterized functions (Bullmore *et al.,* 1996; Kruggel *et al.,* 1999). These benefits open new perspectives for the design of fMRI experiments in cognitive research.

ACKNOWLEDGMENTS

The authors thank D. Norris for providing the datasets and M. Meyer, S. Zysset, and S. Kotz for designing and conducting the fMRI experiments. As a courtesy to the community, we make the testbed including the filter source code available on request.

REFERENCES

Bandettini, P. A., Jesmanowicz, A., van Kylen, J., Birn, R. M., and Hyde, J. S. 1998. Functional MRI of brain activation induced by scanner noise. *Magn. Reson. Med.* **39:**410–416.

Benali, H., Buvat, I., Anton, J. L., Pelegrini, M., Di Paola, M., Bittoun, J., Burnod, Y., and Di Paola, R. 1997. Space-time statistical model for functional MRI image sequences. In *Lecture Notes in Computer Science,* Vol. 1230, *Information Processing in Medical Imaging,* pp. 285–298. Springer, Heidelberg.

Berthod, M., Kato, Z., Yu, S., and Zerubia, J. 1996. Bayesian image classification using Markov random fields. *Image & Vision Comput.* **14:**285–295.

Birn, R. M., Bandettini, P. A., Cox, R. W., Jesmanowicz, A., and Shaker, R. 1998. Magnetic field changes in the human brain due to swallowing or speaking. *Magn. Reson. Med.* **40:**55–60.

Biswal, B., DeYoe, E. A., Jesmanowicz, A., and Hyde, J. S. 1994. Removal of physiological fluctuations from functional MRI signals. In *Proc. SMR, 2nd Annual Meeting, San Francisco,* p. 653.

Biswal, B., DeYoe, E. A., and Hyde, J. S. 1996. Reduction of physiological fluctuations in fMRI using digital filters. *Magn. Reson. Med.* **35:**107–113.

Bullmore, E., Brammer, M., Williams, S. C. R., Rabe-Hesketh, S., Janoth, N., David, A., Mellers, J., Howard, R., and Sham, P. 1996. Statistical methods of estimation and inference for functional MR image analysis. *Magn. Reson. Med.* **35:**261–277.

Buonocore, M. H., and Maddock, R. J. 1997. Noise suppression digital filter for functional magnetic resonance imaging based on image reference data. *Magn. Reson. Med.* **38:**456–469.

Clark, V. P., Maisog, J. M., and Haxby, J. V. 1998. fMRI study of face perception and memory using random stimulus sequences. *J. Neurophysiol.* **79:**3257–3265.

Cohen, M. S. 1997. Parametric analysis of fMRI data using linear systems methods. *NeuroImage* **6:**93–103.

Descombes, X., Kruggel, F., and von Cramon, D. Y. 1998. FMRI signal restoration using an edge preserving spatio-temporal Markov random field. *NeuroImage* **8:**340–349.

Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C., and Evans, A. C. 1994. Assessing the significance of focal activations using their spatial extent. *Hum. Brain Mapp.* **1:**210–220.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., and Turner, R. 1995. Analysis of fMRI time-series revisited. *NeuroImage* **2:**45–53.

Grewal, M. S., and Andrews, A. P. 1993. *Kalman Filtering: Theory and Practice.* Prentice–Hall, Englewood Cliffs, NJ.

Hu, X., and Kim, S. G. 1994. Reduction of signal fluctuation in functional MRI using navigator echoes. *Magn. Reson. Imag.* **31:**495–503.

Hu, X., Le, T. H., Parrish, T., and Erhard, P. 1995. Retrospective estimation and correction of physiological fluctuation in functional MRI. *Magn. Reson. Med.* **35:**201–212.

Jezzard, P., LeBihan, D., Cuenod, C., Pannier, L., Parinster, A., and Turner, R. 1993. An investigation of the contribution of physiological noise in human functional MRI studies at 1.5T and 4T. In *Proc. SMRM, 12th Annual Meeting,* New York.

Kruggel, F., and von Cramon, D. Y. 1999. Modeling the hemodynamic response in single trial fMR experiments. *Magn. Reson. Med.,* in press.

Le, T. H., and Hu, X. 1996. Retrospective estimation and correction of physiological artifacts in fMRI by direct extraction of physiological activity from MR data. *Magn. Reson. Med.* **36:**290–298.

Locascio, J. J., Jennings, P. J., Moore, C. I., and Corkin, S. 1997. Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging. *Hum. Brain Mapp.* **5:**168–193.

Lowe, M. J., and Sorenson, J. A. 1997. Spatially filtering functionally magnetic resonance imaging data. *Magn. Reson. Med.* **37:**723–729.

Lowe, M. J., Mock, B. J., and Sorenson, J. A. 1998. Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuations. *NeuroImage* **7:**119–132.

Noll, D. C., Genovese, C. R., Vazques, A. L., and Eddy, W. 1996. Evaluation of respiratory artifact correction techniques in fMRI using ROC analysis. In *Proc. ISMRM, 4th Annual Meeting,* p. 343.

Petersen, N. V., Jensen, J. L., Burchardt, J., and Støkilde-Jørgensen, H. 1998. State space models for physiological noise in fMRI time series. In *4th Int. Conf. Functional Mapping of the Human Brain,* p. 592.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. 1992. *Numerical Recipes in C,* 2nd ed. Cambridge Univ. Press, Cambridge, UK.

Purdon, P. L., and Weisskoff, R. M. 1998. Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. *Hum. Brain Mapp.* **6:**239–249.

Rorabaugh, C. B. 1993. *Digital Filter Designer's Handbook.* McGraw–Hill, New York.

Scarth, G., Alexander, M., McIntyre, M., Wowk, B., and Somorjai, R. L. 1996. Artifact detection in fMRI using fuzzy clustering. In *Proc. ISMRM, 4th Annual Meeting,* p. 1783.

Sijbers, J., den Dekker, A. J., van Auderkerke, M., Verhoye, M., and van Dyck, D. 1998. Estimation of the noise magnitude in MR images. *Magn. Reson. Imag.* **16:**87–90.

Weisskoff, R. M., Baker, J., Belliveau, J., Davis, T. L., Kwong, K. K., Cohen, M. S., and Rosen, B. R. 1993. Power spectrum analysis of functionally-weighted MR data: What's in the noise. In *Proc. SMRM, 12th Annual Meeting,* New York, p. 7.

West, M., and Harrison, J. 1997. *Bayesian Forecasting and Dynamic Models.* Springer-Verlag, Heidelberg.