NeuroImage **11,** 797–804 (2000)
doi:10.1006/nimg.2000.0558, available online at http://www.idealibrary.com on IDEAL®

Temporal Filtering of Event-Related fMRI Data Using Cross-Validation

Shing-Chung Ngan, Stephen M. LaConte, and Xiaoping Hu

Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Minneapolis, Minnesota 55455

Received October 4, 1999

To circumvent the problem of low signal-to-noise ratio (SNR) in event-related fMRI data, the fMRI experiment is typically designed to consist of repeated presentations of the stimulus and measurements of the response, allowing for subsequent averaging of the resulting data. Due to factors such as time limitation, subject motion, habituation, and fatigue, practical constraints on the number of repetitions exist. Thus, filtering is commonly applied to further improve the SNR of the averaged data. Here, a time-varying filter based on theoretical work by Nowak is employed. This filter operates under the stationary wavelet transform framework and is demonstrated to lead to good estimates of the true signals in simulated data. The utility of the filter is also shown using experimental data obtained with a visual–motor paradigm. © 2000 Academic Press

INTRODUCTION

Since its initial demonstration (Ogawa *et al.*, 1992; Kwong *et al.*, 1992; Bandettini *et al.*, 1992), functional magnetic resonance imaging (fMRI) based on the blood oxygenation level-dependent (BOLD) contrast has emerged as one of the premier tools for studying brain function. Although limited by the hemodynamic response time of the vascular system, temporal resolution of fMRI signals is well within the range of seconds. Recent studies suggest the possibility of pushing this limit further to subsecond range (Menon *et al.*, 1998). Since many cognitive tasks and processes engage the human brain for prolonged periods, time-resolved studies of these activities (also known as event-related fMRI) are desirable. These studies have been shown to be feasible (Buckner *et al.*, 1996; Richter *et al.*, 1997a,b) and represent an exciting avenue for probing neuronal events. At present, however, several technical difficulties in event-related fMRI still pose considerable challenges. One difficulty is the low signal-to-noise ratio (SNR) inherent in the data. Because the BOLD response associated with brain activities is usually very small, the detection and characterization of brain activation signals are nontrivial tasks. As a result, considerable energy within the field of fMRI has been expended on developing new data acquisition and processing techniques to improve the quality of the data.

A common strategy used to improve the SNR in an event-related fMRI study is to collect data with repeated epochs and average the resultant data. The SNR of the averaged epoch is expected to increase as the square root of the number of epochs collected during the experiment. Constraints on the number of repetitions exist, however, because the acquisition time is often limited in practice due to subject motion, habituation, and fatigue. Thus, to improve the SNR of physiological data beyond simple averaging, various data-processing strategies, such as temporal and spatiotemporal filtering, have been adopted. For example, Descombes *et al.* (1998) suggest using Markov random fields in a Bayesian framework to denoise fMRI data in both the space and the time domains simultaneously. Their method exhibits good performance in the detection of activated regions. However, because of the relatively large number of free parameters involved in the algorithm, it is difficult to optimize when the experimental setting changes. Classical temporal filtering techniques such as low-pass filtering suffer from similar problems. The free parameters involved (e.g., the cut-off frequency in the case of low-pass filtering) have to be chosen by the user judicially because inappropriate values will lead to either insufficient noise removal or large distortion of the underlying signal.

In this paper, a temporal filter based upon theoretical work by Nowak (1997) is applied to denoise ensemble-averaged fMRI signals. With their leave-one-out cross-validation approach in the wavelet domain, Nowak *et al.* (1999) have demonstrated the usefulness of the filter in improving the SNR of noisy images obtained from photon-imaging systems. Since there are almost no critical parameters that require adjustment, the filter is easy to use. Here, the effectiveness of the technique in fMRI is demonstrated. In particular, we adapt their filter for the present use by focusing on two aspects. First, the filter will be operating under the stationary wavelet transform (SWT) framework. Using the SWT enhances the denoising performance of the

1053-8119/00 $35.00

resulting filter because it does not suffer from the effect of translation dependence, which is inherent in the standard discrete wavelet transform (Coifman *et al.*, 1995). Second, the performance of filters derived from a more general "leave-*q*-out" framework, where q is less than the total number of epochs K, is studied.

This paper is organized as follows. The basic theories of the SWT and signal estimation using cross-validation are reviewed. Then, a filter drawing from these two theories is described. Subsequently, the effectiveness of the filter in denoising simulated and experimental fMRI signals is demonstrated. Discussion and summary of the results are given at the end of the paper.

THEORETICAL BACKGROUND AND METHODS

Stationary Wavelet Transform

It is well known that discrete wavelet transform (DWT) has the undesirable property of being shift variant (caused by misalignment between features present in the signal and the shape of the wavelet basis elements), leading to instability in the calculation of energy distribution of a given signal across scales. Therefore, the temporal filter to be derived in this paper is based on SWT, which is computationally more expensive but shift invariant. Detailed description of SWT can be found elsewhere (Coifman *et al.*, 1995; Nason *et al.*, 1995; Pesquet *et al.*, 1996). For completeness, the essential ideas presented by Nason *et al.* (1995) are summarized here.

Given a time series $T = \{x_0, x_1, \ldots, x_{N-1}\}$ with $N = 2^J$ for some integer J, the DWT is performed with the combination of a low-pass filter H, a high-pass filter G, and a binary decimation operator D_0. The filter H is defined by a sequence $\{h_n\}_{n\in Z}$, which satisfies the internal orthogonality relation $\sum_n h_n h_{n+2j} = 0$ for any integer j. The filter G is defined as $\{g_n\}_{n\in Z}$, where $g_n = (-1)^n h_{1-n}$. Applying H on each element of the time series T results in $(Hx)_k = \sum_n h_{n-k} x_n$, while applying D_0 on T leads to $(D_0 x)_j = x_{2j}$. Now, the smooth component of the given time series at resolution level J is defined to be the original time series itself,

$$s^J = \{s_n^J\}, \qquad \text{with } s_n^J = x_n \text{ for } n = 0, 1, \ldots, 2^J - 1, \tag{1}$$

and the smooth and the detail components of the time series at resolution level j are defined as

$$s^j = D_0 H s^{j+1} \qquad \text{and} \qquad d^j = D_0 G s^{j+1}, \tag{2}$$

where s^j and d^j are sequences of 2^j terms. Performing the DWT on the time series from level J to some level $L < J$ gives the result

$$Z = \{d^{J-1}, d^{J-2}, \ldots, d^L, s^L\}. \tag{3}$$

Z contains 2^J terms and provides an alternative representation of the given time series.

It can be shown that the binary decimation operator D_0 causes shift variance and hence leads to instability. A remedy to this problem is to apply the high- and low-pass filters to the given time series without decimation. Specifically, we define a new operator Z such that $\{Zx\}_{2j} = x_j$ and $\{Zx\}_{2j+1} = 0$ and define $H^{[r]}$ and $G^{[r]}$ by $\{Z^r h\}$ and $\{Z^r g\}$, respectively. To perform the SWT, set

$$s_n^J = x_n \qquad \text{for } n = 0, 1, \ldots, 2^J - 1 \tag{4}$$

and define the smooth and detail components of the time series at level j as

$$s^{j-1} = H^{[J-j]} s^j \qquad \text{and} \qquad d^{j-1} = G^{[J-j]} s^j, \tag{5}$$

where s^{j-1} and d^{j-1} are sequences of 2^J terms. The collection

$$Z = \{d^{J-1}, d^{J-2}, \ldots, d^L, s^L\} \tag{6}$$

contains $(J - L)2^J$ terms and gives an overdetermined representation of the original time series.

Cross-Validation in Signal Denoising

One basic approach to improve the SNR of physiological data is to fit them with a model of appropriate complexity. For example, in the case of data smoothing by convolution of the raw data with a Gaussian kernel, the model complexity is characterized by the kernel width. Employing an overcomplex model (e.g., a small kernel width) will lead to overfitting and sensitivity to noise inherent in the data, while employing an oversimple model (e.g., a large kernel width) will cause underfitting and large bias in estimating the underlying signal. A number of statistical methods are available to deal with the problem of model selection (Cherkassky *et al.*, 1998). These methods usually involve either (i) derivation or postulation of a penalization function or (ii) resampling. A resampling technique called the "leave-one-out" cross-validation was previously adapted to denoising (Nowak, 1997).

In leave-one-out cross-validation, a given data set of size M is separated into two sets, one of size $M - 1$ (the training set) and the other of size 1 (the validation set). In the framework of event-related fMRI, for example, M would be the total number of epochs. With individual epochs being denoted as $\{g_1(x), g_2(x), \ldots, g_M(x)\}$ and $g_p(x)$ being the epoch belonging to the validation set, one can obtain an estimate $\hat{g}_p(x)$ of the

underlying signal by averaging just the training set, yielding

$$\hat{g}_p(x) = \frac{1}{M-1} \sum_{i \neq p} g_i(x). \tag{7}$$

One can then convolve $\hat{g}_p(x)$, for example, with a Gaussian kernal of width σ to further remove the remaining noise in $\hat{g}_p(x)$,

$$\tilde{g}_p(x) = F(\hat{g}_p(x), \sigma), \tag{8}$$

where F denotes the filtering operation. Now, if the parameter σ is well chosen, $\tilde{g}_p(x)$ will be a good estimate of the underlying uncorrupted signal and in general be a good predictor of the unused epoch $g_p(x)$. In other words, the prediction error in the L_2-norm

$$\epsilon_p(\sigma) = \|F(\hat{g}_p(x), \sigma) - g_p(x)\|^2 \tag{9}$$

should be small. Note that there are M different ways to split the given data set into a training set and a validation set. Hence, an estimate of the expectation of the prediction error reads

$$\epsilon(\sigma) = \frac{1}{M} \sum_{p=1}^{M} \|F(\hat{g}_p(x), \sigma) - g_p(x)\|^2. \tag{10}$$

It is clear that a good value for σ minimizes Eq. (10). Proceeding with this particular choice of σ, the ensemble-averaged epoch can be smoothed. The procedure for parameter selection and denoising is done in the SWT domain and is detailed in the following subsection.

Temporal Filter for Event-Related fMRI Signals

The overall procedure in denoising *ensemble-averaged* fMRI data consists of the following steps. First, the SWT is performed on each epoch of a given time series and also on the ensemble average. Then, wavelet shrinkage is performed separately on each wavelet coefficient obtained from the ensemble average, with the appropriate amount of shrinkage determined by a cross-validation procedure based on dividing the individual epochs into the training and the validation sets. Finally, the result of wavelet shrinkage is transformed back to the time domain, yielding the denoised *ensemble-averaged* fMRI signal.

Before describing the three steps in detail, several notations are defined. Following the terminology used in the beginning of this section, a time series with K epochs, each epoch having N samples, is denoted as

$$\{x_1^0, \ldots, x_1^{N-1}, \ldots, x_k^0, \ldots, x_k^{N-1}, \ldots, x_K^0, \ldots, x_K^{N-1}\}. \tag{11}$$

The ensemble average is

$$\hat{x}^n = \frac{1}{K} \sum_{k=1}^{K} x_k^n. \tag{12}$$

Performing the SWT on an individual epoch k yields

$$\{d_k^{J-1}[0], \ldots, d_k^{J-1}[N-1], d_k^{J-2}[0], \ldots, d_k^{J-2}[N-1], \ldots, d_k^L[0], \ldots, d_k^L[N-1], s_k^L[0], \ldots, s_k^L[N-1]\}. \tag{13}$$

Similarly, the same operation applied to the ensemble average yields

$$\{\hat{d}^{J-1}[0], \ldots, \hat{d}^{J-1}[N-1], \hat{d}^{J-2}[0], \ldots, \hat{d}^{J-2}[N-1], \ldots, \hat{d}^L[0], \ldots, \hat{d}^L[N-1], \hat{s}^L[0], \ldots, \hat{s}^L[N-1]\}. \tag{14}$$

From the wavelet literature, it is known that one can enhance the SNR of a time series by appropriately shrinking its corresponding detail components (i.e., the $d_k^j[n]$) (Donoho, 1995, 1993; Donoho *et al.*, 1994). Here, cross-validation is used to determine the amount of shrinkage needed for each wavelet coefficient in the SWT domain. (See similar treatment in the DWT domain in Nowak (1997).) Analogous to Eq. (7), one can estimate the value for the wavelet coefficient corresponding to the nth bin at resolution level j by

$$\hat{d}_p^j[n] = \frac{1}{K-1} \sum_{k \neq p} d_k^j[n]. \tag{15}$$

Performing the inverse SWT on the coefficients $\hat{d}_p^j[n]$ and $\hat{s}_p^L[n]$ yields an estimate of the underlying signal in the time domain. To denoise this estimate further, each of its wavelet coefficients is shrunk by a factor of $\lambda^j[n]$, where $0 \leq \lambda^j[n] \leq 1$:

$$\tilde{d}_p^j[n] = \hat{d}_p^j[n](1 - \lambda^j[n]). \tag{16}$$

Note that such shrinkage is applied only to the detail and not to the smooth components of the time series. Again, similar to Eq. (10), an estimate of the expectation of the prediction error due to wavelet shrinkage can be written as

$$\epsilon(\Lambda) = \frac{1}{K}\sum_{p=1}^{K}\sum_{n=0}^{N-1}\sum_{j=L}^{J-1}\left\{\frac{1}{2^{J-j}}(\tilde{d}_p^j[n] - d_p^j[n])^2\right\}, \quad (17)$$

where Λ denotes $\{\lambda^{J-1}[0], \ldots, \lambda^{J-1}[N-1], \ldots, \lambda^j[0], \ldots, \lambda^j[N-1], \ldots, \lambda^L[0], \ldots, \lambda^L[N-1]\}$ and the factor $1/2^{J-j}$ normalizes the coefficients in the stationary wavelet domain so that the equation gives a correct value in the L_2-norm. In determining the optimal values for the shrinkage factors in Λ, one solves for $\lambda^j[n]$ in the following equation:

$$\frac{\partial}{\partial \lambda^j[n]}\epsilon(\Lambda) = 0. \quad (18)$$

In the case of a linear model under the leave-one-out cross-validation framework, an analytical solution is possible (Goutte, 1997). In Nowak (1997), such a solution is written in the standard wavelet transform domain. It is clear that the leave-one-out approach for parameter estimation can be generalized to a leave-*q*-out framework. Under this expanded framework, solving an equation analogous to Eq. (18) yields

$$\lambda^j[n] = \gamma\left(\frac{\sum_{\xi\in C(K,q)}(\hat{d}_\xi^j[n])^2 - \sum_{\xi\in C(K,q)} d_\xi^j[n]\hat{d}_\xi^j[n]}{\sum_{\xi\in C(K,q)}(\hat{d}_\xi^j[n])^2}\right), \quad (19)$$

where $\gamma(\theta) = \min(\theta, 1)$ is used to bound the factor $(1 - \lambda^j[n])$ in Eq. (16) between 0 and 1. The set $C(K, q)$ in Eq. (19) refers to the collection of all possible subsets, each of which contains q elements from the list $\{1, \ldots, K\}$; $d_\xi^j[n] = 1/(\#(\xi))\sum_{k\in\xi} d_k^j[n]$ is the average of the *left-out* terms (the validation set), while $\hat{d}_\xi^j[n] = 1/(K - \#(\xi))\sum_{k\in\{1,\ldots,K\}\setminus\xi} d_k^j[n]$ is the average of the *left-in* terms (the training set), with #(ξ) being the number of elements in the set ξ. Now, one can obtain an improved estimate of the true signal by applying wavelet shrinkage on the average of the training set:

$$\tilde{d}_\xi^j[n] = \hat{d}_\xi^j[n](1 - \lambda^j[n]). \quad (20)$$

But applying Eq. (20) to all possible sets ξ in the collection $C(K, q)$ and then taking the average of these results is the same as applying the shrinkage factors directly on the wavelet coefficients corresponding to the *ensemble average:*

$$\tilde{d}^j[n] = \hat{d}^j[n](1 - \lambda^j[n]). \quad (21)$$

Here, $\tilde{d}^j[n]$ denotes the wavelet coefficients for the ensemble average after shrinkage. Finally, an inverse SWT of the coefficients,

$$\{\tilde{d}^{J-1}[0], \ldots, \tilde{d}^{J-1}[N-1], \ldots, \tilde{d}^j[0], \ldots, \tilde{d}^j[N-1], \ldots, \tilde{d}^L[0], \ldots, \tilde{d}^L[N-1], \hat{s}^L[0], \ldots, \hat{s}^L[N-1]\}, \quad (22)$$

gives a *denoised ensemble-averaged* fMRI signal. In essence, the denoising method consists of the successive application of the SWT (Eqs. (13) and (14)), then wavelet shrinkage (Eqs. (19) and (21)), and finally the inverse SWT (Eq. (22)).

Evaluation of the Temporal Filter Using Simulated and Experimental Data

The denoising method detailed in the previous subsection has been implemented in Matlab (The MathWorks, Inc., Natick, MA), with the aid of the WaveLab package (Buckheit *et al.*, 1998). The effectiveness of the method was investigated by applying it to simulated and experimental data. To simulate fMRI data, the following procedure was applied. First, a signal pattern modeling an "idealized" hemodynamic response was generated and replicated as many times as the number of epochs desired. The idealized response was represented with the equation

$$x(t) = \left\{1 - \exp\left(\frac{-t}{\tau_1}\right)\right\}^3 \exp\left(-\frac{t}{\tau_2}\right), \quad (23)$$

where $\tau_1 = 20$ and $\tau_2 = 30$ (time units in TR). Second, noise was superimposed on the simulated signal. The noise used was either Gaussian white noise or baseline data acquired from a healthy volunteer in a resting state (T2*-weighted EPI images, with TR = 300 ms, TE = 55 ms, matrix size 64 × 64, field of view 20 × 20 cm^2). The noise level was scaled according to the SNR desired in a particular numerical experiment, with the SNR defined as the standard deviation of the simulated signal divided by the standard deviation of the noise. To measure the performance of the filter, the ensemble-averaged signal before and after denoising was compared to the uncorrupted signal in terms of the normalized root mean square (NRMS) error. The NRMS error was determined by the L_2-norm between the uncorrupted signal and its estimate, normalized by the standard deviation of the signal.

Simulation

Three numerical experiments were performed to explore the dependence of the algorithm on (1) various levels of SNR, (2) different numbers of epochs, and (3) the choice of mother wavelet. The default settings (unless otherwise mentioned below) for these experiments were as follows. The time courses consisted of 8 epochs, each 64 samples in length. The SNR was 1.0, and a Symmlet wavelet with four vanishing moments (de-

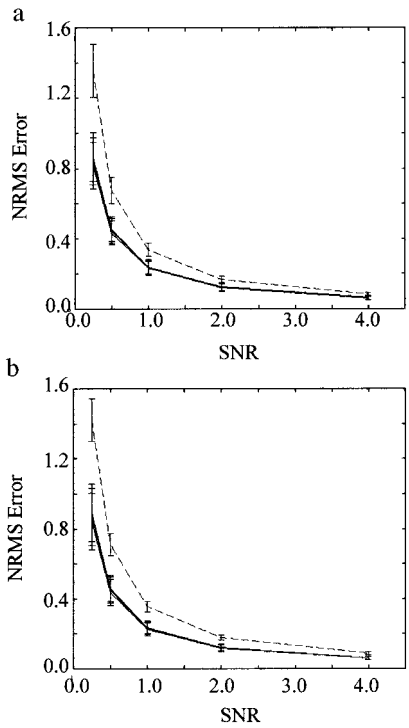

FIG. 1. Performance of the cross-validation filters as a function of the SNR level of the input data. The dashed line refers to signal estimation by the ensemble average, whereas the solid lines refer to signal estimation by the filtered ensemble average. (a) The case in which the true signal is corrupted by the baseline data; (b) the case with Gaussian white noise. The error bars correspond to 1 standard deviation from 10 repeated simulations. The performances of the three cross-validation filters are very close to one another and the corresponding solid lines almost overlap.

noted as Symmlet-4) was employed. For the first experiment, the SNR was varied from 0.25 to 4.0. Both baseline data and Gaussian white noise were used to corrupt the simulated time course. In the second experiment, the number of epochs in each data set ranged from 10 to 30, with each epoch of length 64. Since the baseline data previously obtained were of limited length, only Gaussian white noise was used in this simulation. For the third experiment, the types of mother wavelets studied included Daubechies-4, Daubechies-8, Symmlet-4, and Symmlet-8. In this case, only the baseline data were employed.

Experimental Data

To demonstrate the practical utility of the cross-validation filters, the method was applied to experimental data acquired on a healthy volunteer. The experimental task was to perform rapid finger movement using the dominant right hand when cued with flashing red LED goggles and was repeated 31 times. T2*-weighted EPI images of a single brain slice traversing the occipital cortex and the motor cortex were collected, with TR = 300 ms, TE = 60 ms, matrix size 64 × 64, field of view 22 × 22 cm^2. To quantify the performance of filtering, a "gold" standard was created by averaging all 31 epochs of data. In contrast, only the first 8 epochs of the pixel time courses, representing a shortened version of the experiment, were employed in the calculation and subsequent denoising of the ensemble average. The effectiveness of filtering was gauged by computing the NRMS error of the ensemble average, before and after denoising, relative to the "gold" standard.

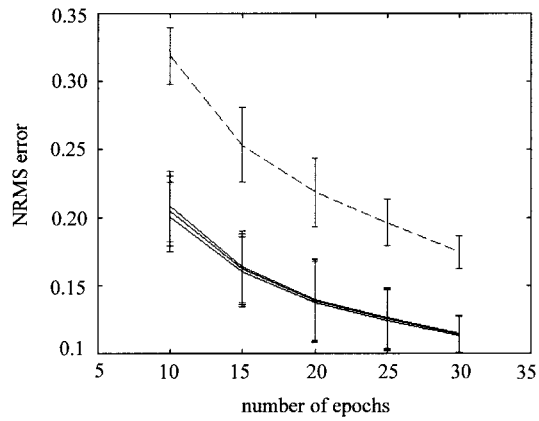

FIG. 2. Filter performance as a function of the total number of epochs in the given time course. The dashed line corresponds to signal estimation by the ensemble average; the solid lines refer to estimation after filtering. The upper solid line is the result from the leave-one-out filter, the middle solid line from the leave-two-out filter, and the bottom solid line from the leave-three-out filter. The error bars correspond to 1 standard deviation.

RESULTS AND DISCUSSION

Performance of the Temporal Filter on Denoising Simulated Data

As shown in Fig. 1, the cross-validation filtering improves the ensemble average at every level of SNR for both the Gaussian white noise and the baseline data

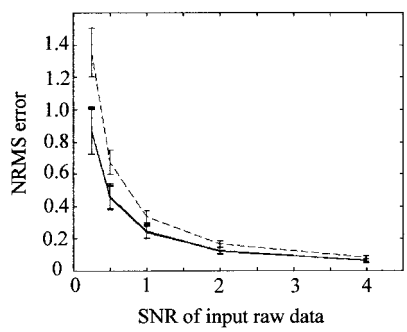

FIG. 3. Filter performance with respect to different types of mother wavelets. The dashed line refers to estimation of the true signal by the ensemble average. The solid lines refer to the result after denoising by the leave-one-out filter. The mother wavelets used are Symmlet-4, Symmlet-8, Daubechies-4, and Daubechies-8. The four lines representing these wavelets are very close to one another.

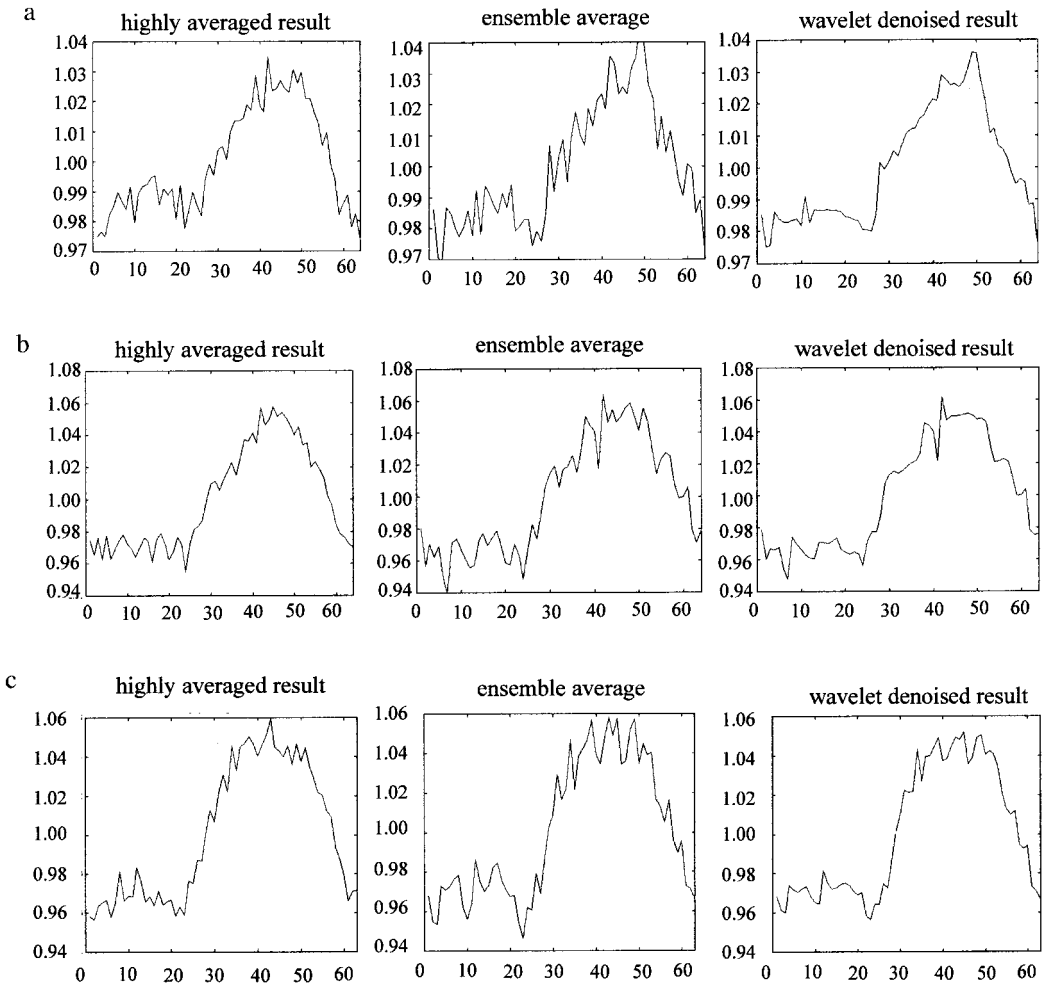

FIG. 4. Filtering of the ensemble-averaged time courses from V1. The leave-one-out filter is used in all the cases here. Three sets of sample pixel time courses are depicted in (a), (b), and (c) separately. The horizontal axis is time and the vertical axis is the normalized signal intensity. The leftmost column is the gold standard (averaging of 31 epochs of data). The middle column is the ensemble average of the first 8 epochs of data. The rightmost column is the new estimate obtained after denoising the middle column.

corrupted signals. Furthermore, the curves corresponding to the filters obtained under the leave-one-out, leave-two-out, and leave-three-out strategies are very close to one another. Close examination of the numerical values (data not shown) reveals that the leave-three-out filter is slightly more effective in denoising than the leave-two-out filter, which in turn is slightly more effective than the leave-one-out filter.

Figure 2 demonstrates that the cross-validation filter reduces the NRMS error of the ensemble average regardless of the number of epochs present in the time course. Inspection of the numerical values reveals that the NRMS error reduction due to filtering is approximately 35–38%. Similar to the first numerical experiment, the leave-three-out filter is slightly more effective than the leave-one-out and the leave-two-out filters in denoising. The increased computational load of the leave-three-out filter compared to the leave-one-out filter, however, is a significant drawback. For example, running the simulation (each simulation with a time course of 25 epochs and each epoch of length 64) 100 times on the SGI O_2 workstation requires 1229, 1321, and 2080 s, respectively, for the leave-one, leave-two, and leave-three cases.

Figure 3 shows that the performance of the cross-validation filter is insensitive to the choice of the mother wavelet. This robustness of the cross-validation filter, which can be attributed to the use of the SWT in place of the standard DWT, implies that the user can freely use one type of wavelet (e.g., Symmlet-4) without concern about possible suboptimality of his or her choice.

Performance of the Temporal Filter on Experimental Data

We selected six sample pixel time courses (three from the V1 region and three from the M1 region) for

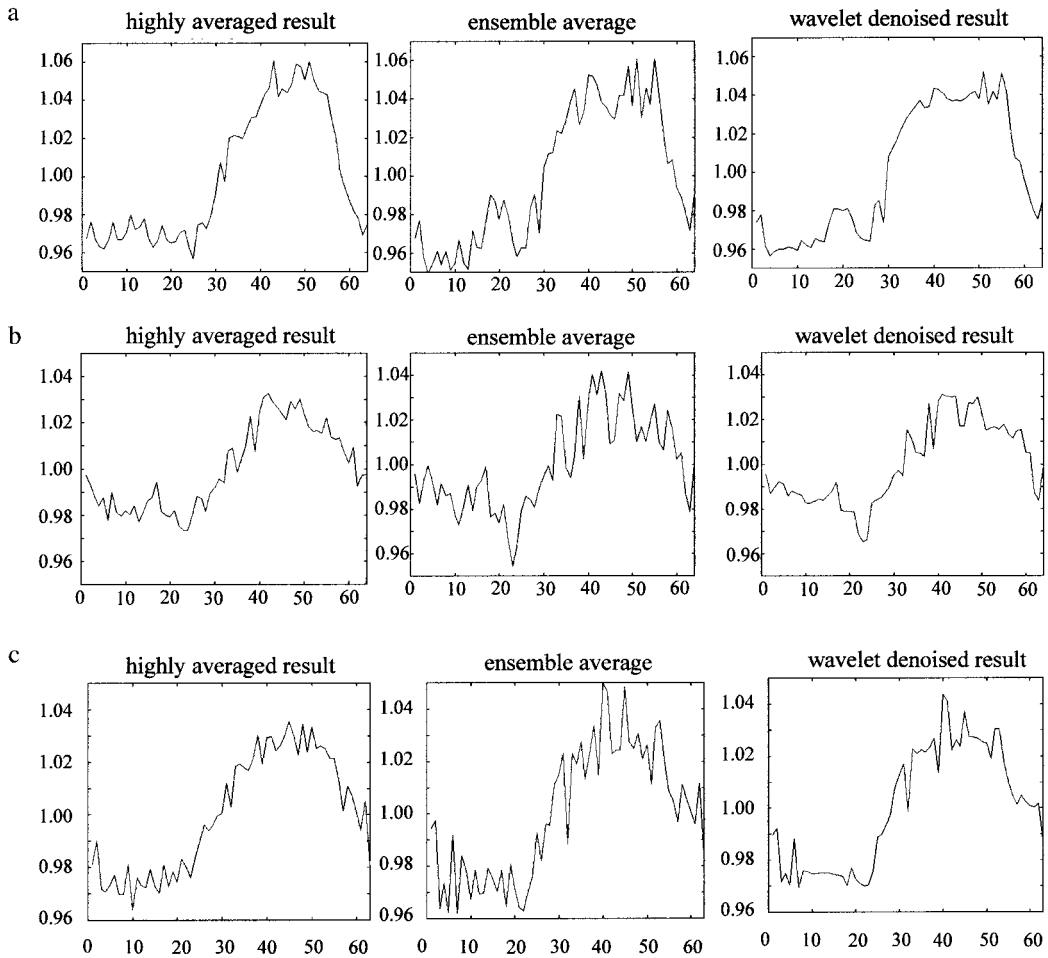

FIG. 5. Filtering of the ensemble-averaged time courses from M1. The leave-one-out filter is used in all the cases here. Three sets of sample pixel time courses are depicted in (a), (b), and (c). The horizontal axis is time and the vertical axis is the normalized signal intensity. The leftmost column is the gold standard (averaging of 31 epochs of data). The middle column is the ensemble average of the first 8 epochs of data. The rightmost column is the new estimate obtained after denoising the time courses from the middle column.

denoising. Figure 4 graphically compares the ensemble-averaged fMRI signals from V1 with their denoised counterparts, showing a visible improvement due to denoising. Similarly, in Fig. 5, the time courses from M1 before and after denoising are compared. Examining the numerical values indicates that the NRMS error reduction is around 18% for the samples from V1 and 29% for the samples in M1. These results confirm the usefulness of the filter in improving the quality of experimental data.

CONCLUSION

In this paper, we adapt theoretical work by Nowak (1997) to construct a time-varying filter for denoising ensemble-averaged fMRI data. The filter is derived from the theory of cross-validation and the SWT. The SWT instead of the standard DWT is utilized because the former does not suffer from the undesirable property of shift variance. To demonstrate the applicability of the resulting filters on fMRI, both simulated and experimental data were used. Numerical analysis and visual inspection of the denoised time courses showed that these filters are effective in improving the signal-to-noise ratio of the raw data without oversmoothing them. A family of these cross-validation filters can be obtained under the general leave-*q*-out framework. In the numerical experiment performed, the performance of the leave-three-out filter is slightly better than that of the leave-two-out framework, which in turn is slightly better than that of the leave-one-out framework. This slight improvement, however, comes with a considerable increase in the computation load. Since the leave-one-out filter is the least computationally expensive among the three, it is recommended for use in fMRI signal denoising.

ACKNOWLEDGMENTS

The authors thank S. Sarkar, E. Yacoub, and J. C. Zhuang for their help in experimental data acquisition, and W. Auffermann for his suggestions on improving the exposition of the manuscript. This

work is supported in part by NIH Grants P41 RR08079, RO1 MH55346, and RO3 MH59245.

REFERENCES

Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., and Hyde, J. S. 1992. Time course EPI of human brain function during task activation. *Magn. Reson. Med.* **25:**390–398.

Buckheit, J., Chen, S., Donoho, D. L., Johnstone, I., and Scargle, J. 1998. WaveLab.701, freely available at http://www-stat.stanford.edu/~wavelab.

Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., and Rosen, B. R. 1996. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA* **93:**14878–14883.

Cherkassky, V., and Mulier, F. 1998. *Learning from Data: Concepts, Theory and Methods*, 1st ed. Wiley, New York.

Coifman, R. R., and Donoho, D. L. 1995. Translation-invariant denoising. In *Wavelets and Statistics* (A. Antoniadis and G. Oppenheim, Eds.). Springer-Verlag, New York.

Descombes, X., Kruggel, F., and Yves von Cramon, D. 1998. Spatio-temporal fMRI analysis using markov random fields. *IEEE Trans. Med. Imag.* **17:**1028–1039.

de Weerd, J. P. C. 1981. A posteriori time-varying filtering of averaged evoked potentials. *Biol. Cybernet.* **41:**211–222.

Donoho, D. L. 1995. Denoising by soft thresholding. *IEEE Trans. IT* **41:**613–627.

Donoho, D. L. 1993. Nonlinear wavelet methods for recovery of signals, densities and spectra from indirect and noisy data. In *Different Perspectives on Wavelets, Proceedings of the Symposium in Applied Mathematics* (I. Daubechies, Ed.), Vol. 47, pp. 173–205. Am. Math. Soc., Providence.

Donoho, D. L., and Johnstone, I. M. 1994. Ideal denoising in an orthonormal basis chosen from a library of bases. Technical Report 461. Department of Statistics, Stanford University, Stanford, CA.

Goutte, C. 1997. *Statistical Learning and Regularisation for Regression*. Ph.D. thesis, University of Paris 6, Paris.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., Cheng, H.-M., Brady, T. J., and Rosen, B. R. 1992. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl. Acad. Sci. USA* **89:**5675–5679.

Menon, R. S., Luknowsky, D. C., and Gati, J. S. 1998. Mental chronometry using latency-resolved functional MRI. *Proc. Natl. Acad. Sci. USA* **95:**10902–10907.

Nason, G. P., and Silverman, B. W. 1995. The stationary wavelet transform and some statistical applications. In *Wavelets and Statistics* (A. Antoniadis and G. Oppenheim, Eds.). Springer-Verlag, New York.

Nowak, R. D. 1997. Optimal signal estimation using cross-validation. *IEEE Signal Process. Lett.* **4:**23–25.

Nowak, R. D., and Baraniuk, R. G. 1999. Wavelet-domain filtering for photon imaging systems. *IEEE Trans. Image Process.* **8:**666–678.

Owaga, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S.-G., Merkle, H., and Ugurbil, K. 1992. Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA* **89:**5951–5955.

Pesquet, J. C., Krim, H., and Carfantan, H. 1996. Time invariant orthonormal wavelet representation. *IEEE Trans. Signal Process.* **44:**1964–1970.

Richter, W., Andersen, P. M., Georgopoulos, A. P., and Kim, S.-G. 1997. Sequential activity in human motor areas during a delayed cued finger movement task studied by time-resolved fMRI. *NeuroReport* **8:**1257–1261.

Richter, W., Andersen, P. M., Georgopoulos, A. P., and Kim, S.-G. 1997. Time-resolved fMRI of mental rotation. *NeuroReport* **8:**3697–3702.