Published in final edited form as:
Neuroimage. 2008 March 1; 40(1): 197–212.

Discriminating brain activity from task-related artifacts in functional MRI: Fractal scaling analysis simulation and application

Jae-Min Lee[1,2], **Jing Hu**[1,2], **Jianbo Gao**[2], **Bruce Crosson**[1,3], **Kyung K. Peck**[4], **Christina E. Wierenga**[1,3], **Keith McGregor**[1,5], **Qun Zhao**[6], and **Keith D. White**[1,2,3,5,*]

1 *Brain Rehabilitation Research Center, Malcom Randall VAMC, Gainesville, FL 32608, USA*

2 *Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA*

3 *Department of Clinical and Health Psychology, University of Florida, Gainesville, FL 32611, USA*

4 *Department of Medical Physics, Memorial Sloan-Kettering Cancer Center, New York, 10021, USA*

5 *Department of Psychology, University of Florida, Gainesville, FL 32611, USA*

6 *Department of Physics Astronomy, University of Georgia Athens, GA 30602, USA*

Abstract

Functional magnetic resonance imaging (fMRI) signal changes can be separated from background noise by various processing algorithms, including the well known deconvolution method. However, discriminating signal changes due to task-related brain activities from those due to task-related head motion or other artifacts correlated in time to the task, has been little addressed. We examine whether three exploratory fractal scaling analyses correctly classify these possibilities by capturing temporal self-similarity; namely, fluctuation analysis, wavelet multi-resolution analysis, and detrended fluctuation analysis (DFA). We specifically evaluate whether these fractal analytic methods can be effective and reliable in discriminating activations from artifacts. DFA is indeed robust for such classification. Brain activation maps derived by DFA are similar, but not identical, to maps derived by deconvolution. Deconvolution explicitly utilizes task timing to extract the signals whereas DFA does not, so these methods reveal somewhat different information from the data. DFA is better than deconvolution for distinguishing fMRI activations from task-related artifacts, although a combination of these approaches is superior to either one taken alone. We also present a method for estimating noise levels in fMRI data, validated with numerical simulations suggesting that Birns model is effective for simulating fMRI signals. Simulations further corroborate that DFA is excellent at discriminating signal changes due to task-related brain activities from those due to task-related artifacts, under a range of conditions.

Introduction

In the past decade, functional magnetic resonance imaging (fMRI) has emerged as a powerful noninvasive tool for studying task-related brain function. The basis for fMRI is blood oxygenation level dependent or BOLD contrast (Ogawa et al., 1993). Oxyhemoglobin, like

*Corresponding author: Department of Psychology, University of Florida, P. O. Box 112250, Gainesville, FL 32611-2250, USA. Tel: (352)392-0601×219, Fax: (352)392-7985, E-mail address: kdwhite@ufl.edu.

water and brain tissue, is weakly diamagnetic (negative susceptibility) while deoxyhemoglobin is paramagnetic (small positive susceptibility). The ratio of deoxy- to oxyhemoglobin in a blood vessel affects the local magnetic field, which in turn affects the precession frequencies of local water protons exposed to the strong main magnetic field used for MRI. Changes in precession frequencies alter the ability to rephase the protons signals, which are spatially encoded by magnetic field gradients. At a place in the brain that is involved in a task, where increased neural activitys metabolic demands convert oxy-to deoxyhemoglobin, there is initially a reduction in BOLD signal strength through changed rephasing. Within a few seconds, an influx of blood oversupplies this locale with fresh oxyhemoglobin, yielding a relatively large signal increase. After the neural activity ends, then the BOLD signal gradually decays toward the pre-activity level. However, it is now well established that there is an extended period of undershoot during which the signal progresses to and remains below prestimulus baseline levels (e.g., Deneux and Faugeras, 2006; Schroeter et al., 2006; Yacoub et al., 2006). This task-related time-varying signal, known as a hemodynamic response (HDR), is the fMRI signature from which underlying neural activity is inferred. The HDR signal amplitude is generally much smaller than the standard deviation of background noise; and because this signal-to-noise ratio is so low, signal detection is uncertain.

A number of signal/image processing algorithms have been developed to separate statistically these task-related signals from noise, including both model-based (Friston et al., 1995; Bullmore et al., 2001; Fadili and Bullmore, 2002; Kamba et al., 2004) and exploratory algorithms (Moser et al., 1997; Backfrieder et al., 1996; Thurner et al., 2003; Shimizu et al., 2004; Friman et al., 2001; McKeown et al., 1998; Muller et al., 2001). Experimentally, a task is repeated many times to gain statistical advantages like time-locked signal averaging in that variations not correlated to the task tend to average out. Unfortunately, causes other than HDRs exist for task-related time-varying RF signals, such as artifacts from movements of the head, mouth, shoulders or arms, or from heart beat and breathing, which produce susceptibility variations within the measurement volume (Birn et al., 1999; Kruger and Glover, 2001; Purdon and Weisskoff, 1998; Diedrichsen and Shadmehr 2005). Such artifacts are quite troublesome when they correlate in time with the very task whose underlying neural activity is the object of study, such as during a study of overt speech production. Not only are task-related artifacts not eliminated by time-locked averaging but they can be wrongly classified as HDRs by some algorithms, giving a misleading picture of the brains activities. Our lab is involved in studies using overt speech and in other studies of patients with movement disorders during which the need to separate HDRs from motion-related signals arises.

Effects of global rigid head motion can be partially mitigated by image registration, most often achieved with intensity-based methods (Cox and Jesmanowicz, 1999; Friston, et al., 1995; Woods, Grafton, Holmes, Cherry, and Mazziotta, 1998), but registration can add artifacts as well as remove them if alignment and intensity interpolations are not accurate (Grootnik, et al., 2000). Even after careful image registration motion can affect the data through spin-history effects (Friston, Williams, Howard, Frackowiak, and Turner, 1996; Bullmore, et al., 1999), changes in the background magnetic field (Jezzard and Clare, 1999), and geometric distortions in echo-planar images (Jezzard and Clare, 1999; Haacke, Brown, and Venkatesan, 1999) or blurring in spiral images (Noll, 1991).

Reduction of motion artifacts has been attempted both prospectively and retrospectively using global rigid motion parameters derived by image registration. Prospective methods include techniques to adjust image acquisition geometry (Thesen, Heid, Mueller, and Schad, 2000; Ward, et al. 2000) or shim field (Ward, Riederer, and Jack, 2002) in quasi-real-time as a function of position in that scan and the one prior to it. Unfortunately, the time between images, albeit only a few seconds, is long enough for the head to move considerably. Retrospective methods include multiple linear regression of the modeled HDR signal and the nuisance signal

estimated by motion parameters from image registration onto fMRI time-series data (Friston, Williams, Howard, Frackowiak, and Turner, 1996; Bullmore, et al., 1999). The drawback with motion parameter regression is that there may not be a linear relation between global rigid motion captured by image registration and local fMRI signal changes, due to the brains magnetic inhomogeneity.

Speech involves local non-rigid movement of pharyngeal muscles, and especially the tongue and jaw, which have been demonstrated to change the magnetic field distribution of the brain (Birn, Bandettini, and Cox, 1998) and lead to significant signal changes in fMRI paradigms that require overt word production (Barch, Sabb, Carter, Braver, and Noll, 1999; Huang, Carr, and Cao, 2001; Crosson, et al., 2005). The act of speaking can induce bulk magnetic susceptibility variations due to changes in airway and vocal tract volumes because oxygen is paramagnetic. Modulation of breathing during speech can also induce task-correlated changes in fMRI signal (Mehta, Grabowski, Razavi, Eaton, and Bolinger, 2006). The greatest magnetic field changes due to speaking occur in the inferior and frontal regions of the brain, decreasing rapidly towards the superior and posterior edges. Most speech-related signal changes occur during the first 4 to 5 seconds after beginning a brief utterance (Birn, Bandettini, and Cox, 1998; Mehta, Grabowski, Razavi, Eaton, and Bolinger, 2006; Birn, Bandettini, and Cox, 1999). HDRs to cognitive processes typically exhibit a delay of up to 6 seconds before onset, and peak 4–6 seconds after that (Cohen, 1997; Hoge, and Pike, 2001). This inherent timing difference between HDR and speech-related artifacts can mitigate the artifacts by ignoring the first few images during and after speech, retrospectively in analysis (Barch, Sabb, Carter, Braver, and Noll, 1999) or prospectively during acquisition (Gracco, Tremblay, and Pike, 2005; Abrahams, et al., 2003; Edmister, Talavage, Ledden, and Weisskoff,. 1999; Hall, et al., 1999). Related approaches are to screen images (voxels) in which artifact is evidenced by temporal changes in the signal phase (Huang, Carr, and Cao, 2001; Soltysik and Hyde., 2006), or as excessively deviant points in the time-series (Huang, Francis, and Carr, 2007).

Ignoring such images does not work well, however, if temporal overlap exists between HDRs and artifacts due to reduced sensitivity for detecting HDRs (Birn, Bandettini, and Cox, 1999; Gopinath, Peck, Soltysik, Crosson, and Briggs, 2003; Gopinath, 2003). The detrending procedure of Birn, Bandettini, and Cox (1999), wherein artifact reduction is achieved by orthogonalizing the time-series to signal changes derived from false positive voxels near the brain edge, also loses sensitivity when HDRs and artifacts overlap (Gopinath, Peck, Soltysik, Crosson, and Briggs, 2003; Gopinath, 2003). Areas involved in response preparation can exhibit HDRs starting before the cue to make the overt response (Cunnington, Windischberger, Deecke, and Moser, 2003) such as supplementary motor area in a word production task (Crosson, et al., 2005; Crosson, et al., 2003; Peck, et al., 2004). Voxels in such regions are particularly susceptible to overlap of HDRs with task-related artifacts, and thus reduced sensitivity for detection of activity using the above approaches.

Motion-related artifacts can be somewhat idiosyncratic and thus tend to average out over a group of participants (Bullmore, et al., 1999; Barch, Sabb, Carter, Braver, and Noll, 1999). The number of subjects needed to ensure that the artifacts vanish in the group averages can be more than 20 in studies conducted in our lab, and this number is difficult to specify in advance. Studies involving intended action may not benefit from this approach because task-related artifacts that are correlated across subjects will not average out. Group averaging also cannot be used when individual differences are meaningful for the brain activity being investigated.

Independent components analysis (ICA) is another form of exploratory data processing that can be useful for classifying and mitigating motion artifacts (Liao, McKeown, and Krolick, 2004). ICA originated for the blind separation of mixed signals (such as a recorded cocktail party) into its sources (individual voices) by deriving an unmixing matrix. As first applied for

fMRI, spatially independent subsets of voxels (components) can be distinguished by the kurtosis of their spatial distributions (McKeown, et al., 1998), and the components can be grouped by cross-correlation of their time series with the task time course. Spatially focal (leptokurtic, super-Gaussian) components were found to be task-correlated whereas spatially dispersed (platykurtic, sub-Gaussian) components found by McKeown, et al (1998) appeared to be motion artifacts not correlated in time with the task. That is, ring-like subsets of voxels near the brain boundary in each slice appeared to be motion artifacts. Complicating this straightforward analysis, unfortunately, are caveats such as that activation changes can be spatially dispersed (e.g., in hypercapnia) as well as that motions can be highly correlated with the task and have relatively local effects (e.g., speech). ICA is useful for extracting information from fMRI (McKeown, Hansen, and Sejnowski, 2003; Calhoun Adali, 2006) but it has not to our knowledge been successfully applied to the problem of separating task-correlated artifacts from HDRs.

In the present paper, we have elected to analyze results from a simple finger tapping task with the aim of separately classifying (a) background noise, (b) task-related HDRs, and (c) artifacts that are task-correlated. Our reasons for selecting this task are discussed further below. Typical fMRI data analyses separate noise (a above) from task-related signals (b and c above mixed together). Specialized analyses for classifying task-related signals (separating b from c) can have the unintended consequence of missing HDRs (mixing a with b). We investigate whether fractal scaling analysis can separate task-related artifacts (c) from task-related HDRs (b). Fractal analysis does not require knowledge of the tasks timing, as do signal averaging or deconvolution methods, as well as do most methods for interpretation of components recovered by spatial ICA, so that fractal analysis might expose somewhat different information about the time series than these other exploratory methods. We also present here an algorithm for estimating noise levels in experimental fMRI signals. Knowing the noise levels in fMRI data can help to establish confidence intervals for activation results.

Fractal characteristics are sometimes useful descriptors of irregular natural phenomena for the purpose of pattern recognition (Gao et al., 2006; Hu et al., 2006). It has been reported that fMRI noise may not be Gaussian but rather fractal (Bullmore et al., 2001; Fadili and Bullmore 2002), and also that BOLD signals without an assigned task are fractal-like (Zarahn et al., 1997). It has furthermore been shown that fractal characteristics of voxel time series discriminate task-related signals from noise using fluctuation analysis (FA; Thurner et al., 2003) or using wavelet multi-resolution analysis (WMA; Shimizu et al., 2004), but it is not presently known whether task-related HDRs can be discriminated from task-related artifacts with either approach. To solve this problem, we presently apply FA, WMA, and a third computational method, novel to fMRI, detrended fluctuation analysis (DFA; Peng et al 1994; Hu et al., 2001; Chen et al., 2002). We study how effectively fractal second order statistics (i.e., temporal self-similarity) distinguish task-related motion artifacts from task-related HDRs, in addition to distinguishing both of these signals from noise.

In order to evaluate sensitivity, specificity, and Receiver Operating Characteristics (ROC) for fractal analyses, we identified a large sample of voxel time series exemplars within human brain images, to serve as the standards for true noise, true task-related artifacts and true task-related HDRs. These exemplars were obtained by deconvolution using Analysis of Functional NeuroImages (AFNI; Cox, 1996). We then evaluated the success of FA, WMA, and DFA for correctly classifying these exemplars from experimental data. We also carried out numerical simulations for the most successful method (DFA), to examine the accuracy of our algorithm for estimating noise level, and to extend the generality of our findings by evaluating the impacts of variations in HDR or motion artifact signal shape, timing, and amplitude.

Our task was visually guided tapping of the right index finger presented in an event-related paradigm. Although one would not expect this task to create artifacts as prominently as if one were speaking aloud, task-related artifacts nevertheless did exist within the data sets of the particular subjects here reported. These particular subjects were selected from a larger cohort (40 participants) because their fMRI images contained more motion artifacts than was true for the majority of the cohort. Instructions to keep still and head packing procedures were apparently less effective for these individuals. It is not particularly uncommon for fMRI studies of normal, healthy subjects using tasks not particularly prone to motion artifacts to exclude the results from a small fraction of the subjects because their images are contaminated by motion artifacts. Some people are not able to remain as still as desired, a problem more frequently encountered in patients with movement disorders, or in some stroke patients attempting simple finger movements who experience overflow. We selected the finger-tapping task for our tests of fractal analysis in part because we work with patients undergoing stroke rehabilitation and with patients having other movement disorders, like Parkinsons disease; unwanted movements are difficult to eliminate in these populations. Prior knowledge of the pattern of brain activations in this task also supports our ability to identify the gold standards for HDRs, artifacts, and noise needed for ROC analysis and for numerical simulations. We believe that one added virtue of separating task-related artifacts from HDRs that took place during this well known task is the capacity to determine whether activation maps derived from our fractal analyses appear credible. This is admittedly a qualitative assessment, but it is one we felt that most investigators would like to evaluate. We furthermore felt that task-related artifacts associated with this finger-tapping task (arising from rigid head motion and out-of-field movements of the shoulders or upper arm) could supply a simpler case for evaluating proof of principle with an untested analytic procedure than would the task-related artifacts arising from overt speech, for the reasons explained in detail above.

Methods

Experimental Participants

All subjects gave written informed consent in accordance with procedures established by the University of Florida Institutional Review Board. Two healthy older male volunteers (age = 71 and 76 years) and two healthy young male volunteers (age = 28 and 29 years) participated. All were strongly righthanded. Exclusion criteria included MRI contraindication; history of neurological disease, dementia or mild cognitive impairment; cardiovascular disease; uncontrolled hypertension; DSM IV Axis 1 diagnosis, learning disability, attention deficit disorder, or substance abuse; and poor visual acuity.

Experimental Imaging Parameters

Functional MRI data were acquired with a 3T Siemens Allegra head-only scanner using gradient-echo echo planar imaging (TR = 1700 ms, TE = 30 ms, 70 deg flip angle, matrix 64 X 64, 240 mm field of view, slice thickness of 5 mm with no gap between slices, and a total of 28 sagittal slices to cover the whole brain). Four runs were scanned, each run consisting of 7 discarded initial images and 83 images obtained subsequently during the event-related behavioral protocol (332 images/session, 28 task events/session).

Task timing was controlled by the viewing of a video monitor. When the participant saw a green star flash 3 times during 1.7 s, they pressed a button three times using the index finger of the right hand as paced by the flashing star. Each such event was followed by a variable interval (i.e., 15.3, 17 or 18.7 s) with only a static red fixation star.

AFNI software (Cox, 1996) reconstructed the DICOM files, which were spatially co-registered with 3dimensional rigid-body transforms. The first 7 images of each run were discarded to

minimize tissue magnetization instability, linear trends within runs were removed, and the runs were then concatenated into time series of 332 images. On a voxel-wise basis, the time series were deconvolved with respect to the time vector for the tasks, to obtain an estimated impulse response function (IRF). How adequately the obtained IRF describes the voxel time series is determined by convolution of the task time vector with the IRF, then fitting this result by least squares regression to the voxel time series, and estimating the proportion of variance explained by that fit (coefficient of determination, R^2).

Selection of exemplar HDRs,task-related artifacts, and noise

Choosing a threshold of $R^2 \geq 0.15$ ($p < 0.001$, uncorrected for multiple comparisons) classified a small fraction of the 128K acquisition voxels as being correlated in time with the task events. We chose this moderately low threshold so as to identify both HDRs and task-related artifacts, the latter of which might not as reliably accompany each task event. Most of the voxels so classified were inside the brain, although a minority of them were outside the brain. We specifically did not ignore, screen, or orthogonalize early time points that might be more likely to contain motion artifacts, as discussed above, because we wished to identify voxels where task-correlated artifacts were significant.

Acquisition voxels outside the body could not be task-correlated on the basis of being HDRs. However, their magnetic signals could instead reflect statistically rare anomalies detected by multiple comparisons, or more interestingly reflect task-correlated artifacts. Similar to Birn, Bandettini, and Cox (1999), we therefore visually examined the estimated IRFs of statistically thresholded voxels outside the brain to exclude the statistical anomalies and to learn the temporal signatures of these artifacts. It is conceivable that task-related signal changes having prolonged temporal signatures could have been incorrectly classified with our procedure as non-task-related statistical anomalies. The majority of statistically thresholded non-brain voxels had abrupt IRF time courses (see Figure 1(d) for an example) but we cannot exclude the possibility that task-correlated artifacts might have protracted time courses. For the present study, the sample of artifact temporal signatures was restricted to relatively abrupt ones, generally lasting under 7 sec, as we and others have found (Birn, Bandettini, and Cox, 1998; Mehta, Grabowski, Razavi, Eaton, and Bolinger, 2006; Birn, Bandettini, and Cox, 1999; Gopinath, Peck, Soltysik, Crosson, and Briggs, 2003; Gopinath, 2003). Armed with a sample of non-brain artifact IRFs, we then chose as exemplars of task-correlated artifacts inside the brain those voxels with IRFs having the same or similar signatures by visual inspection. The remaining task-correlated voxels inside the brain were candidates to be exemplars of HDRs, subject to verification by visual inspection, while brain voxels not meeting the statistical threshold were candidates to be noise. By this combination of statistical thresholding and visual inspection of IRFs, 1800 voxels from the brain images of two healthy volunteers, 600 examples each for HDR, task-related artifact, and noise were identified. We excluded voxels on large veins or at boundaries of the brain, excluded voxels whose IRF appeared to be a mixture of artifact and HDR, and excluded voxels for disagreement between raters on its classification. Voxels were drawn from selected slices in regions of interest related to the motor task, including premotor cortex, supplementary motor area, primary motor cortex, and superior lateral parietal cortex. Slices not used for the selection of exemplars were reserved for subsequent testing of the algorithms generation of credible activation maps. In addition, thousands more noise time series were extracted from the experimental data, by selecting brain areas not expected to be involved in the assigned task (nor detectably related to motion artifacts, heart rate, or breathing) to serve as our empirical noise database for numerical simulation.

Figure 1 shows sample voxel time series (Figs. 1(a,c,e)) for the three kinds of exemplars: HDRs (top), task-related artifacts (middle), and noise (bottom). Task events are indicated in Figures 1(a,c,e) by vertical lines. Adjacent panels (Figs. 1(b,d,f)) are the IRFs obtained by AFNI

deconvolution with respect to the task event vector (Cox, 1996). The noise voxel illustrated was selected from left middle frontal gyrus while the other voxels were selected from left precentral gyrus. Comparing Figs. 1(a,c,e), this HDR's time series has relative peaks soon after many task events, the artifact's time series seems spikier (peaks or valleys near the events), and the noise voxel's time series covered a smaller range of the recorded signal without much evidence of a temporal pattern. The HDR voxel's IRF in Figure 1(b) shows a small initial dip, rise to a peak around the fourth image after the task event (6.8 sec), then a somewhat slower decline a bit below baseline. The artifact's IRF in Figure 1(d) peaks and ends more quickly after the task event. It also begins well below baseline, as though this particular artifact began to appear before the image when the task event took place, perhaps due to anticipation by the subject. Not all artifacts began below baseline, not all were as abrupt as the example shown in Figure 1(d), and some had negative-going rather than positive-going transients. The noise voxel's IRF is unremarkable.

Estimation of noise level from experimental data

fMRI signals are always corrupted by noise. Its presence makes distinguishing HDRs from task-related artifacts difficult; the higher the noise level, the harder the task. To place a confidence interval on the discrimination result, it is important to estimate the level of noise, specifically the noise standard deviation.

We assume (a) that exemplar HDR or artifact time series are the superposition of noise with clean HDR signals or clean artifact signals, respectively, and (b) that these signals and the noise are independent. Thus the variance of an experimental HDR time series $\sigma^2_{\text{noisy-H}}$ is equal to the summation of the variance of the clean HDR signal σ^2_{H} plus the variance of the noise σ^2_{noise}. Like-wise the variance of an experimental artifact time series $\sigma^2_{\text{noisy-a}}$ is equal to the summation of the variance of the clean artifact signal σ^2_{a} plus the variance of the noise σ^2_{noise}. If we knew the $\sigma^2_{\text{noisy-H}}$ or $\sigma^2_{\text{noisy-a}}$ and the σ^2_{noise} then we can easily estimate the variability due to HDR or artifact signals as in Eq. (1).

$$S_{\text{i}} = \sqrt{\frac{\sigma^2_{\text{noisy-i}} - \sigma^2_{\text{noise}}}{\sigma^2_{\text{noise}}}},\ \text{i} = \text{a or H}. \tag{1}$$

Hence, the problem is reduced to finding $\sigma^2_{\text{noisy-H}}$ or $\sigma^2_{\text{noisy-a}}$ and σ^2_{noise}, which we estimate from variances of the exemplar HDRs, artifacts, and noise time series selected earlier. To evaluate the accuracy of our noise level estimation algorithm, we performed numerical simulations using Birn's model for fMRI signals (Birn et al., 2004).

Numerical simulation method

We simulated noise, task-related artifact and HDR time series based essentially on Birn's model (Birn et al., 2004) but with more sources of variability. For example, artifacts can begin at slightly different times than the task events and might vary in amplitude and direction from event to event. HDRs might also vary from one task event to another. Furthermore, the noise residue of fMRI data may not be Gaussian (Bullmore et al., 2001). To insure that our numerical simulations are more representative of actual fMRI data from humans, we therefore used empirically measured fMRI noise rather than generating the noise algorithmically. Details of our simulation procedure follow:

Artifact signal changes are simulated as large spikes in the signal intensity at times coincident with the task event (zero delay) or with one or two TR delay (TR = 1.7 sec in our experiments),

either increasing or decreasing the signal by variable amounts. In Eq. (2), t is the event time, t_s is the delay (0, 1 or 2 TR chosen with equal probability), $\delta(t)$ is the impulse function, and the amplitude parameter a_1 can be either positive or negative with equal probability. Amplitude values were normally distributed with a mean of 3% and a standard deviation of 10% of the baseline. We generated 332 samples per time series (565 sec). These simulated signals approximate the empirical artifact signal changes seen in our motor tasks described earlier.

$$a_1\delta(t - t_s), \tag{2}$$

HDR signal changes were generated by convolving the task event times with a gamma variate function shown in Eq. (3), where the parameter a_2 scales amplitude of the simulated HDR to be 2% of the baseline with a standard deviation of 30% of this amplitude. In addition, we randomly changed the parameters a_3 and a_4 to be uniformly distributed with a means of 8.60 and 0.547 respectively (Cohen, 1997), but with standard deviations of 10% of the corresponding mean. A similar model has been employed in the BOLD signal design for smart phantoms (Zhao et al., 2003).

$$a_2 t^{a_3} e^{-t/a4}, \tag{3}$$

Noise was randomly picked from our noise database and added to the simulated artifact and simulated HDR signal changes. Signal-to-noise ratios (SNRs) were varied by changing the amounts of noise added to those signals.

Computation of fractal statistics

Fractal theory supplies a better description than simple linear mathematics of many irregular objects and phenomena in nature. Mathematically, fractals are characterized by power-law relations (i.e., linear in log-log coordinates) over wide range of scales (Mandelbrot, 1982).

Fluctuation Analysis—Let a voxel time series be denoted as $x(1)$,$x(2)$, …, $x(N)$, where N is a total number of images. FA works as follows. We form the "*random walk*" process $y(n)$, $n = 1, \ldots, N$, by first removing the mean value $\bar{x}$ and then forming partial summation,

$$y(n) = \sum_{i=1}^{n} [\,(x(i) - \bar{x})]. \tag{4}$$

We then examine whether the following scaling law holds or not,

$$F(m) = \sqrt{\langle |y(i+m) - y(i)|^2 \rangle} \sim m^H, \tag{5}$$

where the average is taken over all possible pairs of $(y(i + m), y(i))$. The parameter H is often called the Hurst parameter (Mandelbrot, 1982). When the scaling law described by Eq. (5) holds, then the process under investigation is said to be a fractal process. In fact, when Eq. (5) holds, the autocorrelation for the "*increment*" process, defined as $x(i) = y(i + 1) - y(i)$, decays as a power-law,

$$r(k) \sim k^{2H-2} \; as \; k \to \infty. \tag{6}$$

By the Wiener-Khinchin theorem, the power spectral density for $y(n)$ follows a power-law decay,

$$S_y(f) \sim 1/f^{2H+1}. \tag{7}$$

When $H = 1/2$, the random walk process is similar to standard Brownian motion (Bm), and the increment process is similar to white Gaussian noise (Gn). Generalizations of Bm and Gn are called fractional Brownian motion (fBm) and fractional Gaussian noise (fGn) (Mandelbrot, 1982), characterized by $0 \leq H \leq 1, H \neq 1/2$. When $0 \leq H < 1/2$, the process is said to have anti-persistent correlations, while when $1/2 < H \leq 1$, the process is said to have persistent correlations. Such processes have long memory properties (Mandelbrot, 1982). The latter is justified by noticing that

$$\sum_{k=1}^{\infty} \gamma(k) = \infty. \tag{8}$$

In practice, quite often power-law relations are only valid for a finite region of k. Unfortunately, some researchers try to estimate the H parameter (or other scaling exponents such as the fractal dimension) by some optimization procedure without being concerned about the scaling region.

Wavelet Multi-resolution Analysis—WMA is based on the coefficients of a discrete wavelet de-composition. It involves a scaling function ϕ_0 and a mother wavelet ψ_0. The scaling function satisfies

$$\int_{-\infty}^{\infty} \varphi_0(n) dn = 1.$$

The wavelet ψ_0 must have zero average and decay quickly at both ends (Strang and Nguyen, 1997). The scaled and shifted versions of ϕ_0 and ψ_0 are given by

$$\varphi_{j,k}(n) = 2^{-j/2}\varphi_0(2^{-j}n - k),\ \psi_{j,k}(n) = 2^{-j/2}\psi_0(2^{-j}n - k),\ \ j,k \in Z,$$

where j and k are the scaling (dilation) and the shifting (translation) index, respectively. Different value of j corresponds to analyzing a different resolution level of the signal. One popular technique to perform the discrete wavelet transform is multi-resolution analysis (MRA). The procedure of performing MRA is detailed as follows (Strang and Nguyen, 1997):

1. At the $j = 1$-th resolution, for each $k = 0,1,2, \ldots$, compute the approximation coefficient $a_x(j,k)$ and the detailed coefficient $d_x(j,k)$ according to the following formulae:
$$a_x(j,k) = \sum_n x(n)\varphi_{j,k}(n) = \sum_n x(n)2^{-j/2}\varphi_0(2^{-j}n - k)$$
$$d_x(j,k) = \sum_n x(n)\psi_{j,k}(n) = \sum_n x(n)2^{-j/2}\psi_0(2^{-j}n - k)$$

2. The signal approximation SA_j and the signal detail SD_j at the j-th resolution level are computed as
$$SA_j = \sum_k a_x(j,k)\varphi_{j,k}(n)$$
$$SD_j = \sum_k d_x(j,k)\psi_{j,k}(n)$$

3. Repeat steps (1) and (2) for the $(j + 1)$-th resolution level, using the signal approximation SA_j obtained in step (2) as the input signal.

Let the maximum scale resolution level chosen for analysis be J. The signal can be reconstructed using the following equation (Strang and Nguyen, 1997):

$$x(n) = SA_J + \sum_{j=1}^{J} SD_j = \sum_k a_x(J,k)\varphi_{J,k}(n) + \sum_{j=1}^{J}\sum_k d_x(j,k)\psi_{j,k}(n). \tag{9}$$

The first term represents the approximation at level J, and the second term represents the details at resolution level J and lower. MRA builds a pyramidal structure that requires an iterative application of the scaling and the wavelet functions, respectively. This is schematically shown in Fig. 2.

To make the above procedure more concrete, let us take the Haar wavelet as an example. The scaling function and the mother wavelet of the Haar wavelet are defined as

$$\varphi_0(n) = \begin{cases} 1, & 0 \le n < 1, \\ 0, & \text{elsewhere.} \end{cases}$$
$$\psi_0(n) = \begin{cases} 1, & 0 \le n < 1/2, \\ -1 & 1/2 \le n < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

They are shown in Fig. 3. We consider the signal $x(n)$ consisting of noisy blocks, as shown in Fig. 4(a). The signal approximations and details at resolution levels 1 through 3 are shown in Figs. 4(b,d,f) (left column) and Figs. 4(c,e,g) (right column), respectively. We have

$$x(n) = SA_1 + SD_1 = SA_2 + SD_2 + SD_1 = SA_3 + SD_3 + SD_2 + SD_1.$$

Let

$$\Gamma(j) = \frac{1}{n_j} \sum_{k=1}^{n_j} \left| d_x(j,k) \right|^2,$$

where n_j is the number of coefficients at level j, then the Hurst parameter is given by

$$\log_2 \Gamma(j) = (2H - 1) j + c_0, \tag{10}$$

where c_0 is some constant. When $\log_2 \Gamma(j)$ vs. the scale j curve is approximately linear for certain range of j, the process $x(t)$ is said to be fractal, with slope being $2H - 1$. In particular, a flat horizontal line corresponds to $H = 1/2$.

Detrended Fluctuation Analysis—DFA (Peng et al., 1994; Hu et al., 2001; Chen et al., 2002) is a variant of FA having the distinct advantage that DFA can automatically remove certain trends and/or forms of non-stationarity contained in the time series under study. When applying DFA, one works on the random-walk-type process $y(n)$. DFA works as follows. First, one divides the time series into [N/m] non-overlapping segments (where the notation [x] denotes the largest integer that is not greater than x), each containing m points; then one calculates the local trend in each segment to be the ordinate of a linear least-squares fit for the random walk in that segment, and computes the "*detrended walk*", denoted by $y_m(i)$, as the difference between the original walk $y(i)$ and the local trend; eventually, one examines whether or not the following scaling behavior (i.e., fractal property) holds:

$$F_d(m) = \sqrt{< \sum_{i=1}^{m} y_m(i)^2 >} \sim m^H \tag{11}$$

where the angle brackets denote ensemble average of all the segments and $F_d(m)$ is the average standard deviation over all the segments.

Note that when applying FA or DFA, one works on a random-walk-type process. When one employs WMA, one works on the original time series. For ideal fractal processes, the Hurst parameters estimated by the three methods would be consistent. The Hurst parameter indexes

temporal self-similarity rather than linear correlations in time. It is possible to detect self-similarity in some time series having negligible linear temporal autocorrelation. The length of time the self-similar features last does not specifically predict how the Hurst parameter will be affected. One challenge for using the Hurst parameter as a statistic is that its sampling distribution cannot be derived analytically. The sampling distribution can, however, be approximated empirically through permutations of the data, either by shuffling them in the time domain or by randomizing phases in the Fourier domain.

Results and Discussion

Experimental Results

Figure 5 shows representative fluctuation analysis (FA) outcomes for each type of exemplars, HDRs (top), task-related artifacts (middle), and noise (bottom). The slopes of lines fitted to scaling ranges in these log-log plots estimate the Hurst parameter, a fractal second-order statistic. We observed that HDRs have the largest H (steepest slope on the log-log plot), followed by task-related artifacts, then noise. Our findings agree with those of Thurner et al. (2003), whose fluctuation analysis showed that HDRs have larger H values than do noise. From Figures 5(a, b), we observe that fractal scaling for HDRs breaks down around $m = 2^{2.5}$, which corresponds to a time scale of around 10 sec (5 to 6 TRs), also agreeing with Turner *et al.* (2003). By contrast, Figs. 5(c)–(f) show that noise and task-related artifact time series have a wider fractal scaling range (~ 20 s).

Figure 6 shows outcomes from wavelet multi-resolution analysis (WMA) for each type of exemplars, HDRs (top), task-related artifacts (middle) and noise (bottom). As in Figure 5, HDRs have the largest H (steepest slope on the log-log plot), followed by task-related artifacts, then noise. H values obtained by WMA are always larger than corresponding ones obtained by FA. From Figs. 6(a, b), we observe that fractal scaling for HDRs breaks down at $j = 3$. By contrast, Figs. 6(c)–(f) show that noise and task-related artifact time series have a wider fractal scaling range.

Figure 7 shows outcomes from DFA for each type of exemplars, HDRs (top), task-related artifacts (middle) and noise (bottom). As in Figs. 5 and 6, HDRs have the largest H (steepest slope), followed by task-related artifacts, then noise. H values obtained by DFA are always larger than corresponding ones obtained by FA. Also, the breakdown of fractal scaling happens at longer times for DFA than for FA. This is particularly clear for HDRs, for which scaling now extends nearly three-fold to $m = 2^4$ (~ 27 sec, 16 TRs).

The fMRI time series for task-related artifacts and for noise are very well defined by power law relations (i.e., linear relations in log-log plots) over a moderately wide range of scales. Scale ranges obtained using DFA are much longer than those obtained using FA. The fMRI data may be non-stationary or have trends that FA cannot remove as does conventional power spectral density analysis (Koscielny-Bunde, 1998). On the other hand, DFA can automatically remove certain trends and non-stationarity (Peng et al., 1994; Hu et al., 2001; Chen et al., 2002). The fractal scaling behavior for HDRs identified by DFA is longer than our 20 sec average inter-stimulus interval, suggesting that DFA is capturing the undershoot phase of the HDR better than FA, WMA, or deconvolution (with its selected parameters).

To explore the effectiveness of H for classifying voxel time series, we identified by R^2 thresholding and visual inspection of IRFs, 1800 voxels from brain images of two healthy volunteers, 600 examples each for HDR, task-related artifact, and noise. We then applied FA, WMA, and DFA to all 1800 voxel time series. Consistent with the findings above and with Thurner *et al.* (2003), we observed that the H values for noise are always smaller than those

for HDRs, whether *H* was calculated by FA, WMA, or DFA. Thus it was quite easy to separate noise from HDRs.

The more challenging task was to distinguish task-related artifacts from HDRs. We carried out Receiver Operation Characteristic (ROC) analysis (Hanley and McNeil, 1982), a commonly used method for summarizing the relation between sensitivity and specificity of a measure, to evaluate the discrimination performance of *H* as calculated by FA, WMA, and DFA. Figures 8(a,b,c) plot sensitivity and specificity of *H* for discriminating 600 HDRs from 600 task-related artifacts, comparing *H* as calculated by FA, WMA, or DFA, respectively. The solid lines plot sensitivity (1 -probability of false negative) and the broken lines plot specificity (1 - probability of false positive). Figure 8(d) shows the corresponding ROC curves. Optimal discrimination maximizes the probability of true positives, while minimizing the probability of false positives. DFA clearly provides better performance than FA and WMA.

An optimal decision rule that minimizes the maximum of two error probabilities (minimax approach) makes the sensitivity and specificity of *H* equal. In Figs. 8(a,b,c) each vertical line shows the respective minimax-optimized thresholds for *H*, 0.59 for FA (panel a), 0.73 for WMA (panel b) and 0.78 for DFA (panel c). With these thresholds the corresponding probabilities of true positive are 0.75 for FA, 0.81 for WMA and 0.91 for DFA, shown by horizontal lines in Figs. 8(a,b,c). Alternatively, the Neyman-Pearson (N-P) criterion is to choose a threshold under the constraint of controlling the probability of false positives (Kay, 1998). Constraining the probability of false positives to 0.05 results in N-P-optimized thresholds for $H = 0.67$ for FA, $H = 0.84$ for WMA, and $H = 0.80$ for DFA. With these latter thresholds the corresponding probabilities of true positive are 0.45 for FA, 0.52 for WMA and 0.81 for DFA. For DFA, the likelihood ratio of true to false positives is slightly over 16 to one.

Figure 9 shows three slices of the brain as "activation maps" created by applying FA (panel a), WMA (panel b), DFA (panel c), or deconvolution with R^2 thresholding (panel d). Those maps were created with FA threshold $H = 0.67$, WMA threshold $H = 0.84$ and DFA threshold $H = 0.80$, corresponding to the N-P optimization $P_{FA} = 0.05$, and threshold $R^2 = 0.15$ ($p < 0.001$, uncorrected for multiple comparisons). These particular slices were not used during the identification of exemplars for the computations of sensitivity, specificity, and ROC curves. For our experimental paradigm, which involved right-hand finger tapping and attended visual stimulation, activations are expected in the hand bump (Yousry, 1997) of left primary motor cortex (M1) and in the primary visual cortex (V1). FA (panel a) and WMA (panel b) unfortunately find relatively little activation in M1 or V1. FA and WMA instead identify a smattering of locations, often single voxels, laterally and medially over the frontal lobes. We do not find FA and WMA activation maps to be credible. Activation maps defined by FA (panel a) or by WMA (panel b) differ from the maps defined by DFA (panel c) or by R^2 thresholding (panel d), however. The expected M1 and V1 activations are indeed observed in DFA and R^2 maps, lending them credibility. DFA and R^2 activation maps are fairly similar but not identical. R^2 thresholding identifies some small regions, many being individual voxels, that DFA does not. DFA identifies as active a substantial subpopulation of the voxels also declared active by R^2 thresholding, whereas FA and WMA do not. Since we have found that FA and WMA are similar for our purpose, to simplify presentation below we shall only compare results obtained using FA and DFA.

The 116 voxels to be used in Fig. 10 are those identified by deconvolution with $R^2 \geq 0.15$ in the three slices presented in Fig. 9. These were classified by visual inspection of their estimated impulse response functions into 75 HDRs and 41 task-related artifacts, by the procedures given in Methods. Figure 10 plots these HDRs as open circles and these artifacts as crosses to compare deconvolution's R^2 statistic to *H*, the latter calculated either by FA (left panel) or by DFA (right panel). Horizontal lines indicate the N-P criterion thresholds for *H* given 0.05 probability of

false positives. FA does a poor job of separating HDRs (circles) from artifacts (crosses). Specifically, while FA yields few false positives, it also yields few true positives (HDRs). On the other hand, DFA is roughly 87% successful (4 artifacts classified as HDRs, 11 HDRs classified as artifacts, neither of which is significantly different from the ROC analysis expected error rates). Thus, DFA seems to strike a good balance between minimizing false positives and false negatives, whereas FA could not produce such a result even by shifting the cutoff for H. Higher values of R^2 are more likely to be HDRs than artifacts, and vice versa for low values of R^2, but HDRs (circles) and artifacts (crosses) have more overlap along the R^2 axis than they do along the H axis obtained by DFA.

To further elucidate activations defined by R^2 and DFA methods, we identified the 111 voxels with DFA $H \geq 0.80$ (likely HDRs) in the three slices presented in Fig. 9. A subset of 68 of these also had $R^2 \geq 0.15$, thus were deemed active by both methods (H+R+). The remaining 43 deemed as likely HDRs by DFA H were deemed to be noise by R^2 thresholding (H+R−). Lastly, of the 116 voxels deemed active by R^2, 37 voxels were deemed to be task-related artifacts rather than HDRs by DFA (H−R+). Figure 11 plots the average event-related time courses (IRFs) associated with H+R+, H+R−, and H−R+ subsets of voxels, in panels a, b, and c respectively. The vertical bars indicate the estimated standard errors of the mean (expressed as $\sigma/\sqrt{N}$, where σ is the standard deviation and N is the sample size). We observed the most consistent hemodynamic response results when both DFA and deconvolution agree about the voxel being active (panel a). That is, most time points on the average IRF exclude zero from their confidence intervals. This seems like evidence that more information is extracted by combining DFA with deconvolution than by using either analysis alone. The question is, for the two subsets of voxels where the methods disagree, does the average IRF time course resemble the HDR in one case but not the other? When the analyses disagree about voxel activity, DFA is more consistent on average (panel b) than is deconvolution (panel c) for excluding the null.

Most brain activation maps involve choosing a threshold for the test statistics. Sometimes activation patterns change dramatically with slight variation of the choice of threshold, which reduces one's confidence in the outcomes. We therefore created additional activation maps (not shown) from another subject for the following ranges of test statistic thresholds: $H = 0.66$ to 0.72 for FA, $H = 0.83$ to 0.92 for DFA, and $R^2 = 0.18$ to 0.24. We concluded that the FA maps, like those shown in Fig. 9 top row, did not reveal consistent activations in M1, V1, or other logically possible areas like pre-motor cortex or supplementary motor areas. DFA and R^2 maps showed such activations consistently, despite varying their thresholds over this range of decision criteria.

Estimation of noise from experimental data

We have prepared three databases by selecting 1800 voxels from the brain images of two healthy volunteers, 600 examples each for HDRs, task-related artifacts, and noise. Within each database, we calculate the average variance among all the time series. The estimated variance for true HDRs, task-related artifacts, and noise is about 1.52, 1.37, and 0.99, respectively. By Eq. (1), we obtain $S_H = 0.73$ and $S_a = 0.62$.

To evaluate the accuracy of the estimations of the noise level, we performed the following numerical simulations. We simulated 5000 HDR time series using Eq. (3) with $S_H = 0.73$ and 5000 task-related artifact time series using Eq. (2) with $S_a = 0.62$. We then applied DFA to the random-walk-type processes of all these simulated voxel time series, computed the ROC curve of H, and compared it with the ROC curves of H and R^2 for real data. These three curves were shown in Fig. 12 as dashed line, solid line, and dash-dot line, respectively. They are very close. Note that if we increase S_H and S_a by decreasing the noise level, the dashed line will move up.

Conversely, if we decrease S_H and S_a by increasing the noise level, the dashed line will move down. This suggests that the proposed algorithm for estimating noise level is very effective. In fact, Fig. 12 also suggests that Birn's model (Birn et al., 2004) is a good model for fMRI signals, and corroborates that DFA is excellent at distinguishing true HDRs from our sample of task-related artifacts.

Conclusions

We compared fluctuation analysis (FA; Thurner et al., 2003), wavelet multi-resolution analysis (WMA; Shimizu et al., 2004), and detrended fluctuation analysis (DFA) for characterizing fMRI data as though these time series were random fractals having temporal self-similarity. We used Receiver Operating Characteristic analysis to examine the effectiveness of FA, WMA, and DFA in distinguishing among three types of fMRI time series, namely hemodynamic responses (HDRs), task-related artifacts, and background noise. Since it is easy to separate the noise voxels from HDRs by various methods, we focused instead on distinguishing HDRs from task-related artifacts, because such artifacts can comprise a significant proportion of a brain activation map (Birn et al., 1999; Krüger and Glover, 2001; Purdon and Weisskoff, 1998; Diedrichsen and Shadmehr 2005). Nearly without regard to the choice of decision criteria for separating HDRs from artifacts, the likelihood ratio of true positives to false positives is substantially larger for DFA than for FA and for WMA. Activation maps derived by DFA are similar to those derived by deconvolution analysis with R^2 thresholding. R^2 thresholding separates task-related signals from noise but is not particularly effective for separating task-related HDRs from task-related artifacts. However, DFA separates task-related HDRs from task-related artifacts in the present finger-tapping paradigm, as well as separating both from noise.

We have also proposed an effective method for estimating the noise level in experimental fMRI data, and we have partially evaluated through simulations the effects of varying lag, amplitude, and time course shape on the Receiver Operating Characteristics for DFA and deconvolution. We emphasize that this evaluation is only partial. Whether DFA will perform as well for the non-rigid motion artifacts that accompany overt speech, or for motion-related signal changes having time courses more similar to those of HDRs, is not yet known.

Deconvolution analysis explicitly uses information about task timing to extract the activation patterns while DFA does not. Conceivably, HDRs not paced by a task might be detectable by DFA if they happened with sufficient density within the time series. Deconvolution and DFA expose somewhat different information about the time series. It seems logical to speculate that one could effectively integrate these two methods, as crudely illustrated in Figure 11, to further improve both sensitivity and specificity for detecting functional activity of brain areas involved in tasks. It seems logical to speculate as well that DFA might be useful for categorizing components recovered through spatial ICA.

Finally, other investigators are having success using nonlinear techniques to characterize BOLD responses. Deneux and Faugeras (2006) showed that maximum likelihood parameter estimation could be used with a physiological model to derive activation maps. Future studies should compare such methods with those described or developed in the present paper to assess the performance of the different methods for classifying different kinds of artifacts under varying constraints.

Acknowledgements

This work was supported by Brain Rehabilitation Research Center VA Center of Excellence grant VARR&D F2182C, Research Career Scientist Award VARR&D B3470S to BC, NIH grant P50-DC03888 (BC, core PI), and the Evelyn F. McKnight Brain Research Grant Program at the University of Florida (JBG, KDW). The present paper is an extension of a conference presentation at the 2005 IEEE International Conference on Acoustics, Speech, and Signal

Processing, “Identification of brain activity by fractal scaling analysis of functional MRI data,” authored by J.M. Lee, J. Hu, J.B. Gao, K.D. White, B. Crosson, C. Wierenga, K. McGregor, and K.K. Peck.). We thank Kaundinya Gopinath, Department of Radiology, Univeristy of Texas Southwestern Medical Center at Dallas, for useful discussions.

References

Abrahams S, Goldstein LH, Simmons A, Brammer MJ, Williams SC, Giampietro VP, Andrew CM, Leigh PN. Functional magnetic resonance imaging of verbal fluency and confrontation naming using compressed image acquisition to permit overt responses. Hum Brain Map 2003;20:29–40.

Backfrieder W, Baumgartner R, Samal M, Moser E, Bergmann H. Quantification of intensity variations in FMRI using rotated principal components. Phys Med Biol 1996;41:1425–1438. [PubMed: 8858728]

Barch DM, Sabb FW, Carter CS, Braver TS, Noll DC. Overt verbal responding during fMRI scanning: empirical investigations of problems and potential solutions. NeuroImage 1999;10:642–657. [PubMed: 10600410]

Basho S, Palmer ED, Rubio MA, Wulfeck B, Mller RA. Effects of generation mode in fMRI adaptations of semantic fluency: paced production and overt speech. Neuropsychologia 2007;45:1697–1706. [PubMed: 17292926]

Birn RM, Bandettini PA, Cox RW, Jesmanowicz A, Shaker R. Magnetic field changes in the human brain due to swallowing or speaking. Magn Res Med 1998;40:55–60.

Birn RM, Bandettini PA, Cox RW, Shaker R. Event-related fMRI of tasks involving brief motion. Hum Brain Map 1999;7:106–114.

Birn RM, Cox RW, Bandettini PA. Experimental designs and processing strategies for fMRI studies involving overt verbal responses. NeuroImage 2004;23:1046–1058. [PubMed: 15528105]

Bullmore E, Long C, Suckling J, Fadili J, Calvert G, Zelaya F, Carpenter TA, Brammer M. Colored noise and computational inference in neurophysiological (fMRI) time series analysis: Resampling methods in time and wavelet domains. Hum Brain Map 2001;12:61–78.

Bullmore ET, Brammer MJ, Rebe-Hesketh S, Curtis VA, Morris RG, Williams SCR, Sharma T, McGuire PK. Methods for diagnosis and treatment of stimulus-correlated motion in generic brain activation studies using fMRI. Hum Brain Map 1999;7:38–48.

Calhoun VD, Adali T. Unmixing fMRI with independent component analysis. IEEE Eng Med Biol 2006:79–90.

Chen Z, Ivanov PC, Hu K, Stanley HE. Effect of nonstationarities on detrended fluctuation analysis. Phys Rev E 2002;65:Art No 041107 Part 1.

Cohen MS. Parametric analysis of fMRI data using linear system methods. NeuroImage 1997;6:93–103. [PubMed: 9299383]

Cox RW. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimage. Comput Biomed Res 1996;29:162–173. [PubMed: 8812068]

Cox RW, Jesmanowicz A. Real-time image registration for functional MRI. Magn Res Med 1999;42:1014–1018.

Crosson B, Benefield H, Cato MA, Sadek JR, Moore AB, Wierenga CE, Gopinath K, Soltysik D, Bauer RM, Auerbach EJ, Gkay D, Leonard CM, Briggs RW. Left and right basal ganglia and frontal activity during language generation: contributions to lexical, semantic, and phonological processes. J Intl Neuropsychol Soc 2003;9:1061–1077.

Crosson B, McGregor K, Gopinath K, Conway TW, Benjamin M, Chang Y, Moore AB, Raymer AM, Briggs RW, Sherod MG, Wierenga CE, White KD. Functional MRI of language in aphasia: A review of the literature and the methodological challenges. Neuropsychol Rev 2007;17:157–177. [PubMed: 17525865]

Crosson B, Moore AB, Gopinath K, White KD, Wierenga CE, Gaiefsky ME, Fabrizio KS, Peck KK, Soltysik D, Milsted C, Briggs RW, Conway TW, Gonzalez Rothi LJ. Role of the right and left hemispheres in recovery of function during treatment of intention in aphasia. J Cogn Neurosci 2005;17:392–406. [PubMed: 15814000]

Cunnington R, Windischberger C, Deecke L, Moser E. The preparation and readiness for voluntary movement: a high-field event-related fMRI study of the Bereitschafts-BOLD response. NeuroImage 2003;1:404–412. [PubMed: 14527600]

Deneux T, Faugeras O. Using nonlinear models in fMRI data analysis: model selection and activation detection. NeuroImage 2006;32:1669–1689. [PubMed: 16844388]

Desco M, Hernandez JA, Santos A, Brammer M. Multiresolution analysis in fMRI: sensitivity and specificity in the detection of brain activation. Hum Brain Map 2001;14:16–27.

Diedrichsen J, Shadmehr R. Detecting and adjusting for artifacts in fMRI time series data. NeuroImage 2005;27:624–634. [PubMed: 15975828]

Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM. Improved auditory cortex imaging using clustered volume acquisitions. Hum Brain Map 1999;7:89–97.

Fadili MJ, Bullmore ET. Wavelet-generalized least squares: a new BLU estimator of linear regression models with 1/f errors. NeuroImage 2002;15:217–232. [PubMed: 11771991]

Friman O, Cedefamn J, Lundberg P, Borga H, Knutsson H. Detection of neural activity in functional MRI using canonical correlation analysis. Magn Reson Med 2001;45:323–330. [PubMed: 11180440]

Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frackowiak RSJ. Spatial registration and normalization of images. Hum Brain Map 1995;3:165–189.

Friston KJ, Holmes AP, Worseley KJ, Poline JB, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Map 1995;2:189–210.

Friston KJ, Williams S, Howard R, Frackowiak RSJ, Turner R. Movement-related effects in fMRI time-series. Magn Res Med 1996;35:346–355.

Gao JB, Hu J, Tung WW, Cao YH, Sarshar N, Roychowdhury VP. Assessment of long range correlation in time series: How to avoid pitfalls. Phys Rev E 2006;73:016117.

Gopinath, KS. PhD Thesis. University of Florida; Gainesville, FL USA: 2003. Reduction of noise due to task correlated motion in event related overt word generation functional magnetic resonance imaging paradigms.

Gopinath KS, Peck KK, Soltysik DA, Crosson BA, Briggs RW. A selective detrending method to reduce noise due to event-correlated motion in fMRI time-series for an event-related overt word generation paradigm. Proc Intl Soc Magn Res Med 2003;11:388.

Gracco VL, Tremblay P, Pike B. Imaging speech production using fMRI. NeuroImage 2005;26:294–301. [PubMed: 15862230]

Grootnik S, Hutton C, Ashburner J, Howseman AM, Josephs O, Rees G, Friston KJ, Turner R. Characterization and correction of interpolation effects in the realignment of fMRI time-series. NeuroImage 2000;11:49–57. [PubMed: 10686116]

Haacke, EM.; Brown, RW.; Venkatesan, R. Magnetic resonance imaging: Physical principles and sequence design. New York: Wiley; 19999.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. Sparse temporal sampling in auditory fMRI. Hum Brain Map 1999;7:213–223.

Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29–36. [PubMed: 7063747]

Heim S, Amunts K, Mohlberg H, Wilms M, Friederici AD. Head motion during overt language production in functional magnetic resonance imaging. Neuroreport 2006;17:579–582. [PubMed: 16603915]

Hoge, RD.; Pike, GB. Quantitative measurement using fMRI. In: Jezzard, P.; Matthews, PM.; Smith, SM., editors. Functional MRI: An introduction to methods. 1. Oxford: Oxford University Press; 2001. p. 159-174.

Hu J, Gao JB, Posner FL, Zheng Y, Tung WW. Target detection within sea clutter: a comparative study by fractal scaling analyses. Fractal 2006;14:187–204.

Hu K, Ivanov PC, Chen Z, Carpena P, Stanley HE. Effect of trends on detrended fluctuation analysis. Phys Rev E 2001;64:Art No 011114 Part 1.

Huang J, Carr T, Cao Y. Comparing cortical activations for silent and overt speech using event-related fMRI. Hum Brain Map 2001;15:39–53.

Huang J, Francis AP, Carr TH. Studying overt word reading and speech production with event-related fMRI: A method for detecting, assessing, and correcting articulation-induced signal changes and for measuring onset time and duration of articulation. Brain Lang. 2007in press

Jezzard P, Clare S. Sources of distortion in functional MRI data. Hum Brain Map 1999;8:80–85.

Kamba M, Sung YW, Ogawa S. A dynamic system model-based technique for functional MRI data analysis. NeuroImage 2004;22:179–187. [PubMed: 15110008]

Kay, SM. Fundamentals of Statistical Signal Processing: Detection Theory. Prentice Hall; New Jersey: 1998.

Koscielny-Bunde E, Bunde A, Havlin S, Roman HE, Golderich Y, Schellnhuber H-J. Indication of a Universal Persistence Law Governing Atmospheric Variability. Phys Rev lett 1998;81:729–732.

Krüger Gand Glover GH. Physiological noise in oxygenation-sensitive magnetic resonance imaging. Magn Res Med 2001;46:631–637.

Liao R, McKeown MJ, Kroli JL. Isolation and minimization of effects of motion on fMRI using multiple reference images. Proc IEEE Intl Symp Biomed Imag: Nano to Macro 2004;1:916–919.

Mandelbrot, BB. The Fractal Geometry of Nature. W.H. Freeman; New York: 1982.

McKeown MJ, Makeig S, Brown GG, Jung TP, Kindermann SS, Bell AJ, Sejnowski TJ. Analysis of fMRI data by blind separation into independent spatial components. Human Brain Mapping 1998;6:160–188. [PubMed: 9673671]

Mehta S, Grabowski TJ, Razavi M, Eaton B, Bolinger L. Analysis of speech-related variance in rapid event-related fMRI using a time-aware acquisition system. NeuroImage 2006;29:1278–1293. [PubMed: 16412665]

Moser E, Diemling M, Baumgartner R. Fuzzy clustering of gradient-echo function MRI in the human visual cortex. Part II: Quantification. Magn Reson Imaging 1997;7:1094–1101.

Muller K, Lohmann G, Bosch V, von Cramon DY. On multivariate spectral analysis of fMRI time series. NeuroImage 2001;14:347–356. [PubMed: 11467908]

Ogawa S, Menon RS, Tank DW, Kim SG, Merkle H, Ellermann JM, Ugurbil K. Functional brain mapping by blood oxygenation level-dependent contrast magnetic-resonance-imaging - a Comparison of signal characteristics with a biophysical model. Biophysical Journal 1993;64:803–812. [PubMed: 8386018]

Peck KK, Moore AB, Crosson BA, Gaiefsky M, Gopinath KS, White KD, Briggs RW. Functional magnetic resonance imaging before and after aphasia therapy: shifts in hemodynamic time to peak during an overt language task. Stroke 2004;35:554–559. [PubMed: 14739418]

Peng CK, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL. Mosaic organization of DNA nucleotides. Phys Rev E 1994;49:1685–1689.

Purdon PL, Weisskoff RM. Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. Hum Brain Map 1998;6:239–249.

Schroeter ML, Kupka T, Mildner T, Uludag K, von Cramon DY. Investigating the post-stimulus undershoot of the BOLD signal - a simultaneous fMRI and fNIRS study. NeuroImage 2006;30:349–358. [PubMed: 16257236]

Shimizu Y, Barth M, Windischberger C, Moser E, Thurner S. Wavelet-based multi-fractal analysis of fMRI time series. NeuroImage 2004;22:1195–1202. [PubMed: 15219591]

Skudlarski P, Constable RT, Gore JC. ROC analysis of statistical methods used in functional MRI: individual subjects. NeuroImage 1999;9:311–329. [PubMed: 10075901]

Soltysik DA, Hyde JS. Strategies for block-design fMRI experiments during task-related motion of structures of the oral cavity. NeuroImage 2006;29:1260–1271. [PubMed: 16275020]

Strang Gand Nguyen, T. Wavelet and filter banks. Wellesley-Cambridge Press; New York: 1997.

Thesen S, Heid O, Mueller E, Schad LR. Prospective acquisition correction for head motion with image-based tracking for real-time fMRI. Mag Res Med 2000;44:457–465.

Thurner S, Windischberger C, Moser E, Walla P, Barth M. Scaling laws and persistence in human brain activity. Physica A 2003;326:511–521.

Ward HA, Riederer SJ, Grim RC, Ehman RL, Felmlee JP, Jack CR. Prospective multi-axial motion correction for fMRI. Mag Res Med 2000;43:459–469.

Ward HA, Riederer SJ, Jack CR. Real-time autoshimming for echo planar time-course imaging. Mag Res Med 2002;48:771–780. 521.

Woods RP, Grafton ST, Holmes CJ, Cherry SR, Mazziotta JC. Automated image registration: I. General methods and intrasubject, intramodality validation. J Comput Assist Tomogr 1998;22:139–152. [PubMed: 9448779]

Yacoub E, Ugurbil K, Harel N. The spatial dependence of the poststimulus undershoot as revealed by high-resolution BOLD- and CBV-weighted fMRI. J Cereb Blood Flow Metab 2006;26:634–644. [PubMed: 16222242]

Yousry TA, Schmid UD, Alkadhi H, Schmidt D, Peraud A, Buettner A, Winkler P. Localization of the motor hand area to a knob on the precentral gyrus – A new landmark. Brain 1997;120:141–157. [PubMed: 9055804]

Zarahn E, Aguirre GK, DEsposito M. Empirical analyses of BOLD fMRI statistics: I. Spatially unsmoothed data collected under null-hypothesis conditions. NeuroImage 1997;5:179–197. [PubMed: 9345548]

Zhao Q, Duensing GR, Fitzsimmons J. Smart Phantom for Characterizing, Validating and Comparing Neuro Informatics Tools. Proceedings of the 12nd Annual Meeting of the International Society of Magnetic Resonance in Medicine 2003;11:1834.

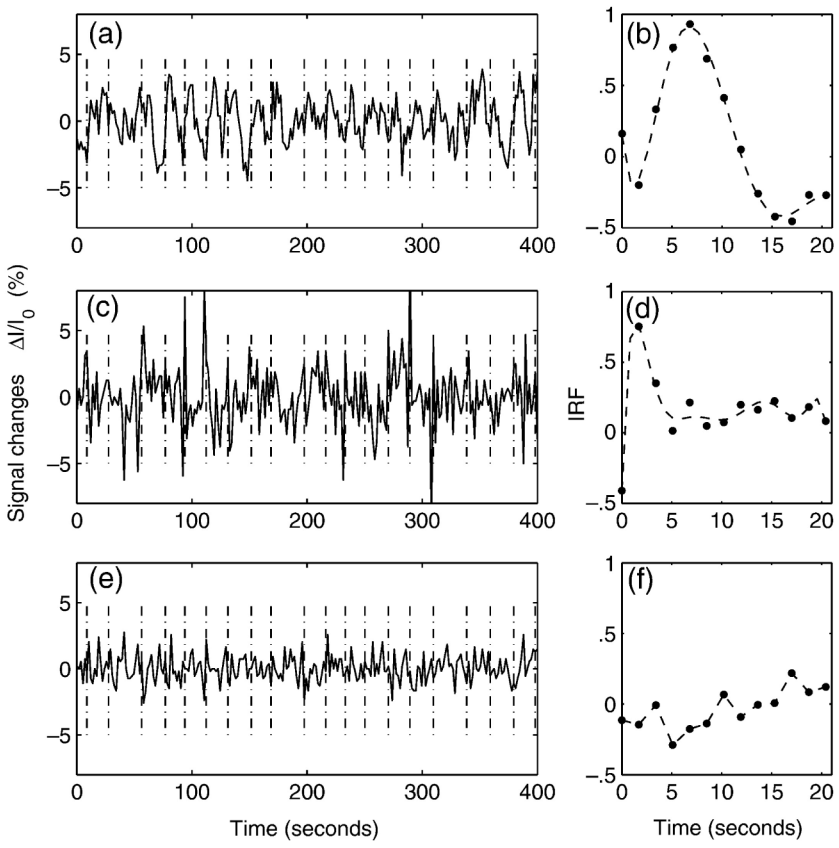

Figure 1.

Representative time series and the corresponding estimated impulse response functions for: (a, b) hemodynamic response, (c, d) task-related artifact, and (e, f) background noise. Vertical lines in (a, c, e) indicate the times of task events. Smooth curves in (b, d) represent the estimated impulse response functions fitted by 7th order polynomials.

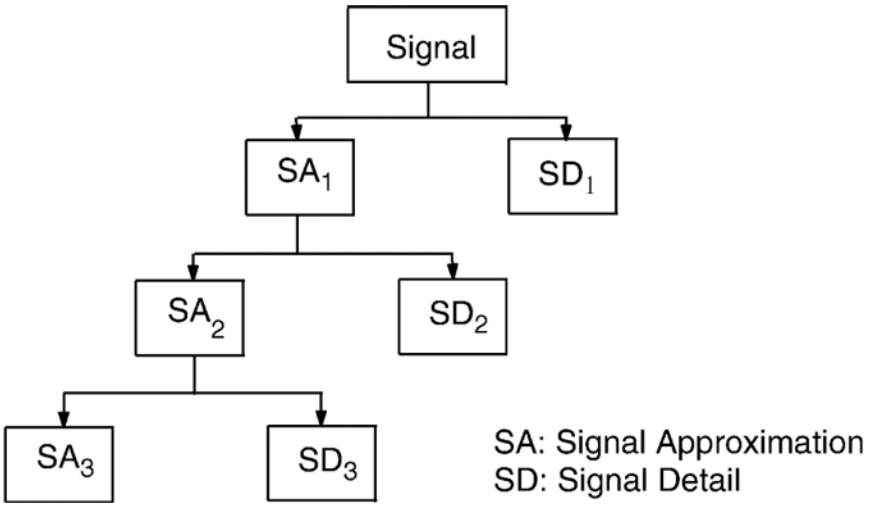

Figure 2.
Pyramidal structure of the output of wavelet multi-resolution analysis.

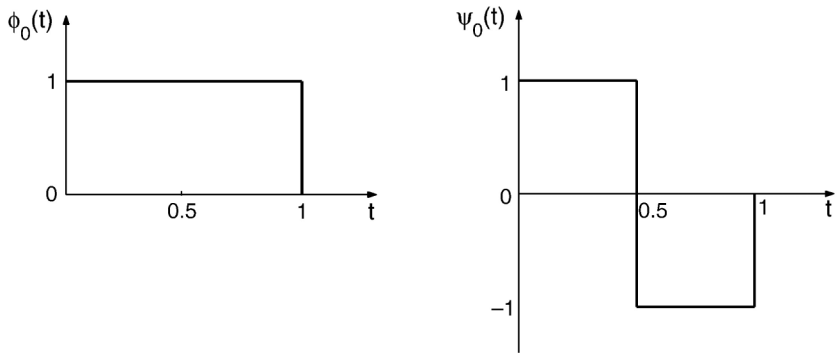

Figure 3.
The scaling function $\phi_0(n)$ and the mother wavelet $\psi_0(n)$ of the Haar wavelet.

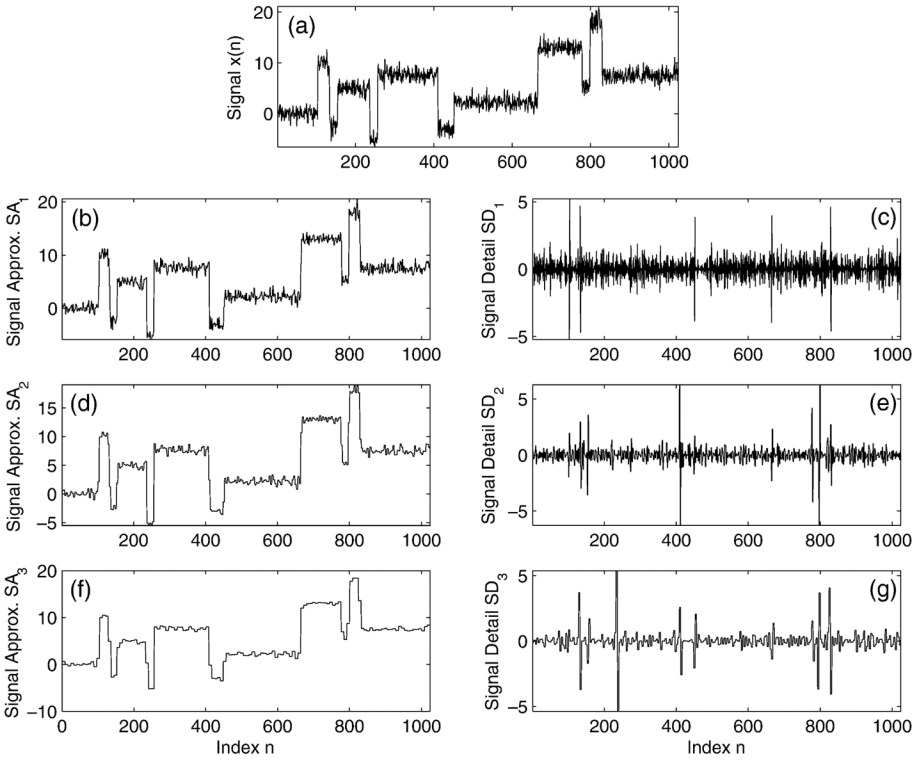

Figure 4.

(a) The input signals $x(n)$, (b,d,f) and (c,e,g) are the signal approximations and the signal details at resolution levels 1 through 3, respectively. $x(n) = SA_1 + SD_1 = SA_2 + SD_2 + SD_1 = SA_3 + SD_3 + SD_2 + SD_1$.

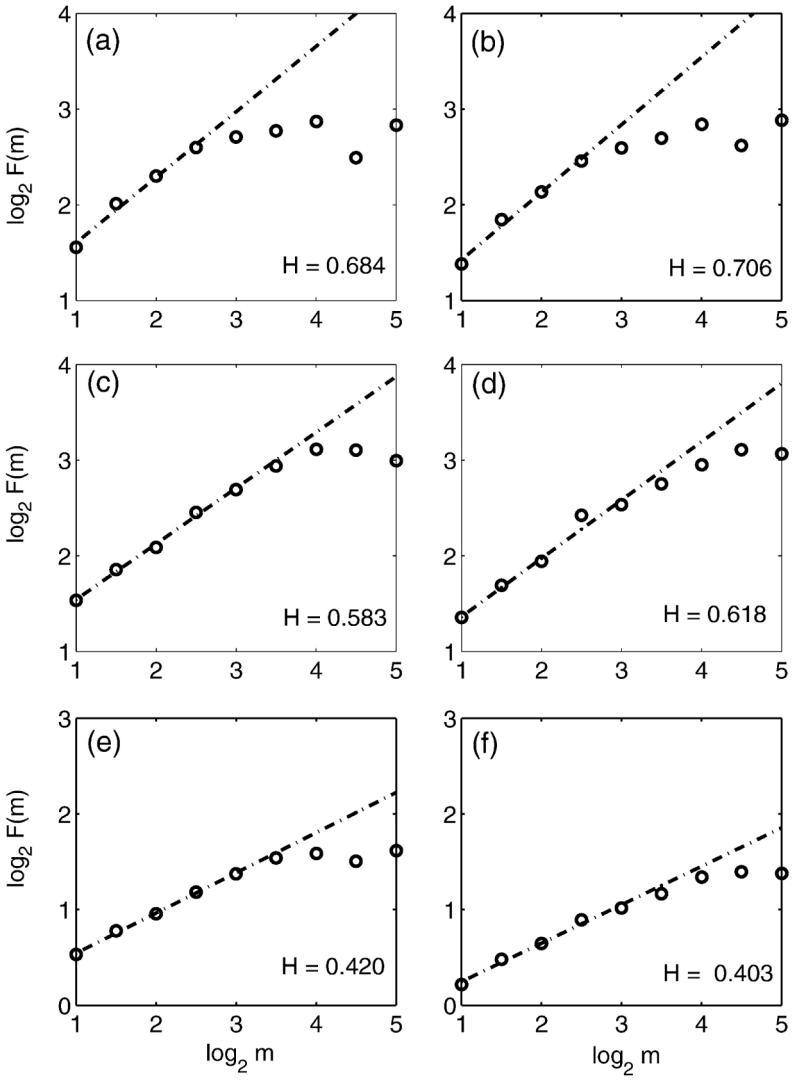

Figure 5.
Fluctuation analysis for the three types of voxel time series. (a, b) hemodynamic responses, (c, d) task-related artifacts, and (e, f) background noise.

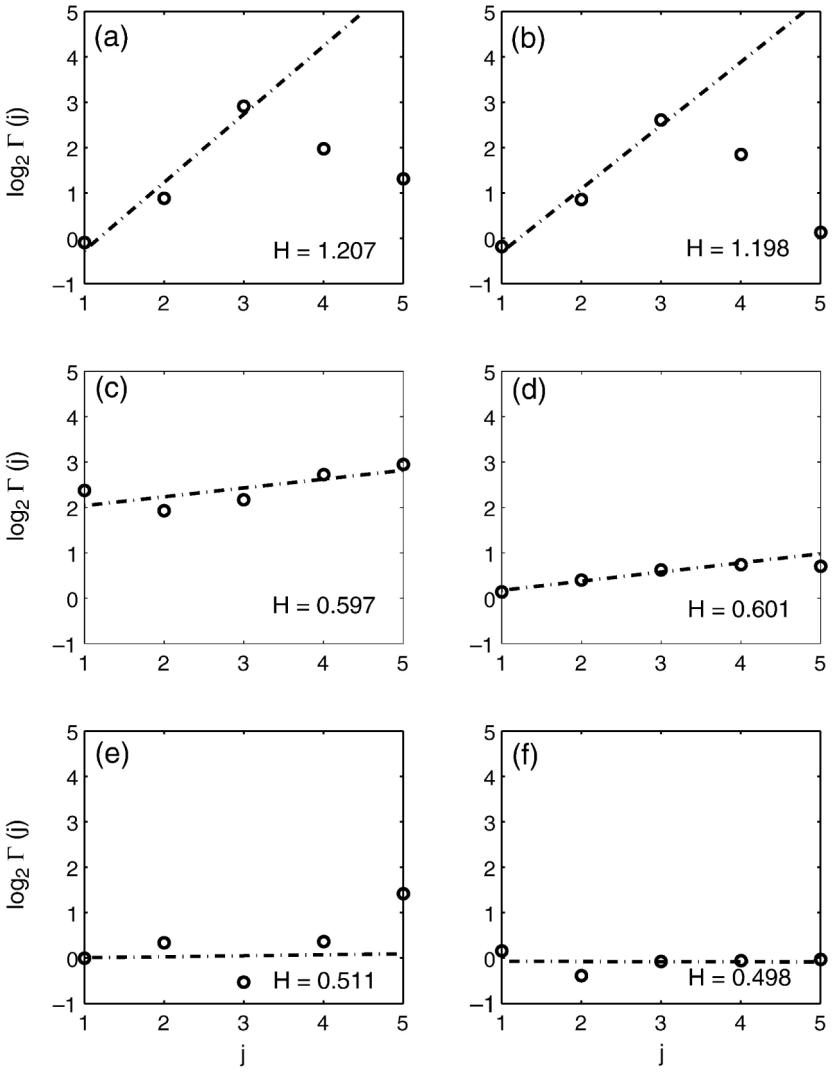

Figure 6.
Wavelet multi-resolution analysis for the three types of voxel time series. (a, b) hemodynamic responses, (c, d) task-related artifacts, and (e, f) background noise.

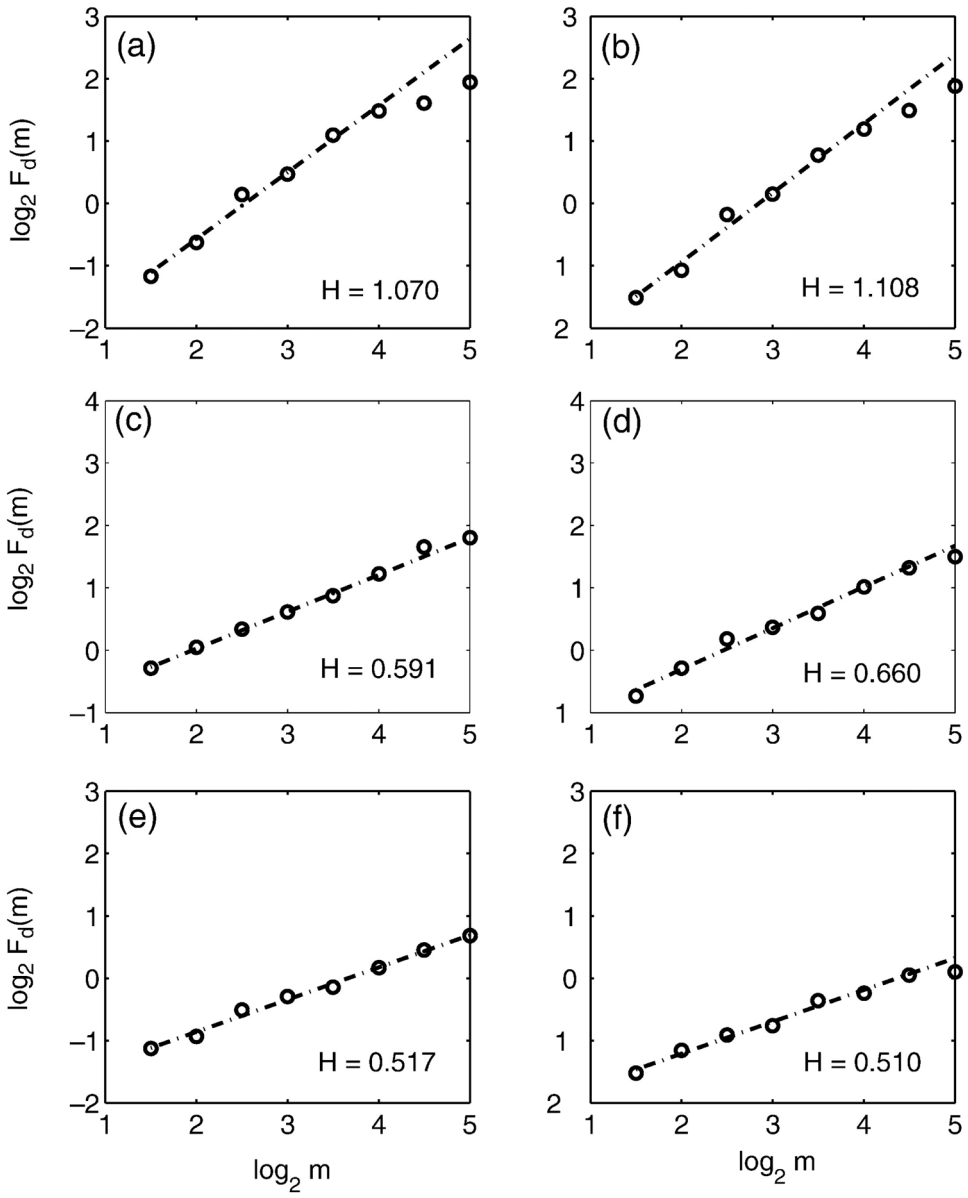

Figure 7.
Detrended fluctuation analysis for the three types of voxel time series. (a, b) hemodynamic responses, (c, d) task-related artifacts, and (e, f) background noise.

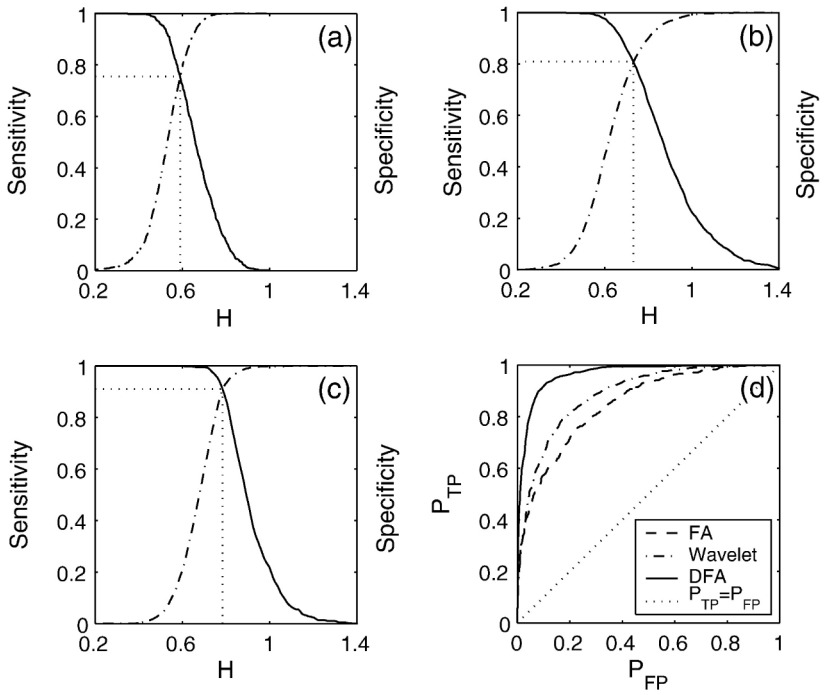

Figure 8.

Sensitivity and specificity for *H* to discriminate hemodynamic responses from task-related artifacts, as calculated by fluctuation analysis (panel a), by wavelet multi-resolution analysis (panel b), and by detrended fluctuation analysis (panel c). Vertical lines indicate values of *H* for which sensitivity is equal to specificity. Panel d shows the corresponding Receiver Operating Characteristic curves.

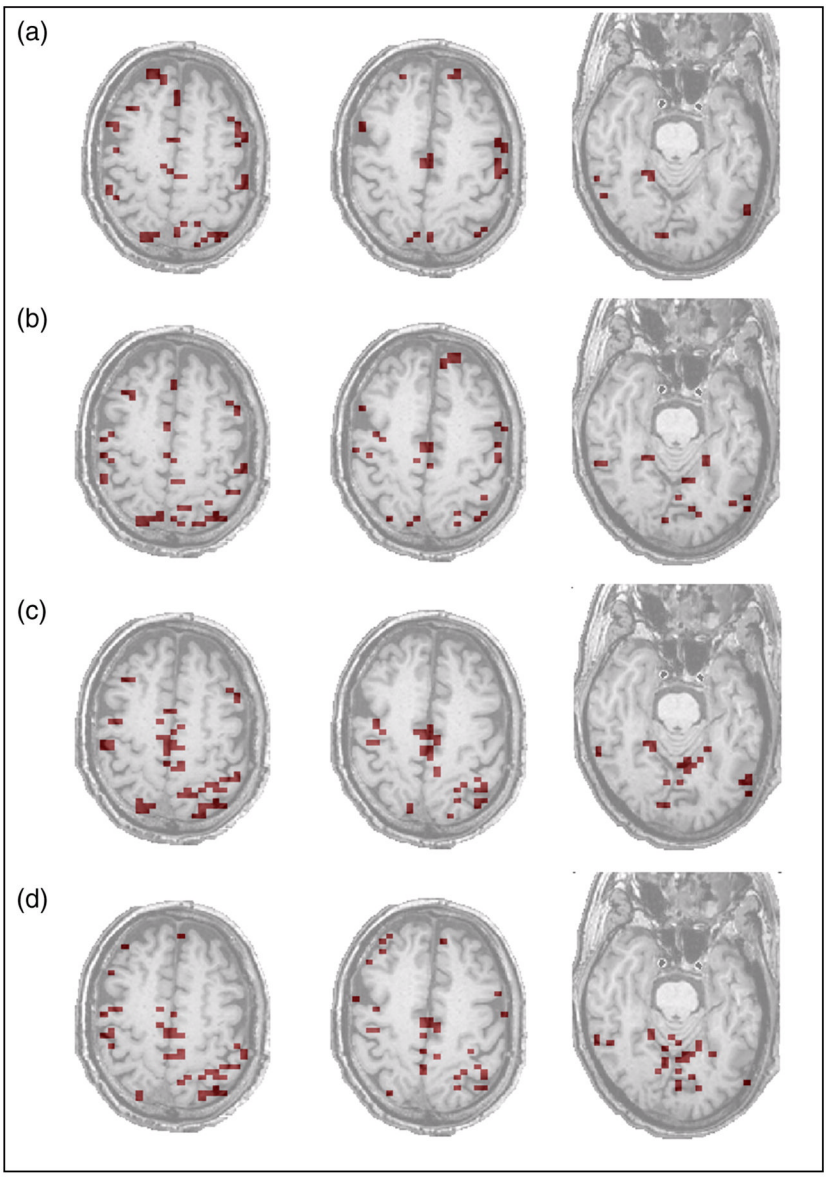

Figure 9.

Brain activation maps of three representative axial slices. (Panel a) Fluctuation analysis thresholded at $H = 0.67$, probability of false positive $P_{FP} = 0.05$. (Panel b) Wavelet multi-resolution analysis thresholded at $H = 0.84$, probability of false positive $P_{FP} = 0.05$. (Panel c) Detrended fluctuation analysis thresholded at $H = 0.80$, probability of false positive $P_{FP} = 0.05$. (Panel d) Results of deconvolution thresholded at $R^2 = 0.16$. All brain maps are oriented with the right hemisphere shown on the left.

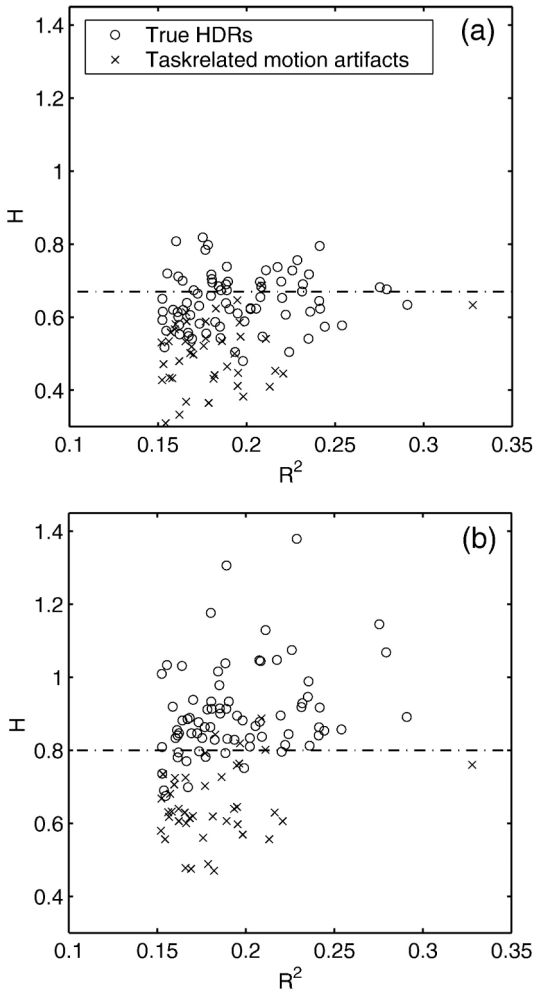

Figure 10.

Comparison of R^2 statistics and H parameters for suprathreshold voxels (n = 116) in the three slices shown in Figure 9(d). These voxels were subdivided by inspection of the impulse response functions into 75 hemodynamic responses (circles) and 41 task-related artifacts (crosses). (Panel a) Fluctuation analysis thresholded at H = 0.67 does not clearly separate hemodynamic responses from artifacts. (Panel b) Detrended fluctuation analysis thresholded at H = 0.80 does separate them.

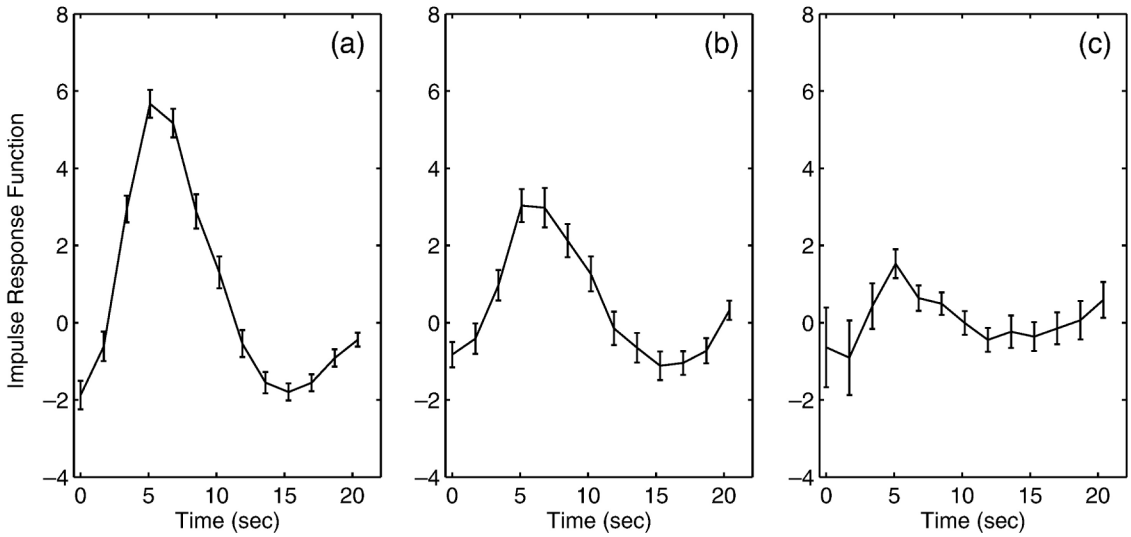

Figure 11.

Average event-related time courses associated with: (Panel a) Voxels declared active by both Hurst parameter and R^2 statistical thresholds (H+R+, 68 voxels); (Panel b) Voxels declared active by Hurst parameter but not R^2 statistical thresholds (H+R−, 43 voxels); (Panel c) Voxels not declared active by Hurst parameter but declared active by R^2 statistical thresholds (H−R+, 37 voxels). The vertical bars indicate the estimated standard errors of the means.

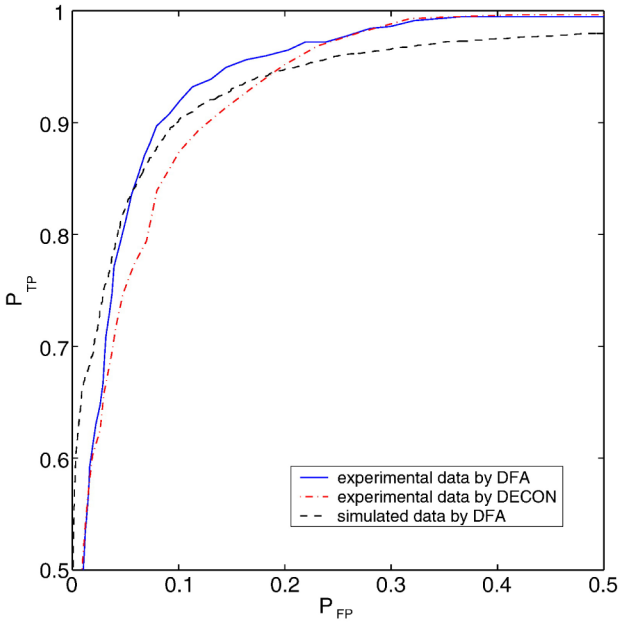

Figure 12.

The solid, dash-dot, and dashed lines denote the ROC curves for experimental data using DFA, experimental data using deconvolution, and simulated data of similar SNR (S_H = 0.73 and S_a = 0.62) using DFA, respectively.