Selective Detrending Method for Reducing Task-Correlated Motion Artifact During Speech in Event-Related FMRI

Kaundinya Gopinath,[1]* **Bruce Crosson,**[2,3] **Keith McGregor,**[4] **Kyung Peck,**[5] **Yu-Ling Chang,**[4] **Anna Moore,**[3,6,7] **Megan Sherod,**[2] **Christy Cavanagh,**[2] **Ashley Wabnitz,**[2] **Christina Wierenga,**[2,3] **Keith White,**[4] **Sergey Cheshkov,**[1] **Venkatagiri Krishnamurthy,**[1] **and Richard W. Briggs**[1]

[1]*Department of Radiology, UT Southwestern Medical Center, Dallas, Texas*
[2]*Department of Clinical and Health Psychology, University of Florida, Gainesville, Florida*
[3]*VA RR&D Brain Rehabilitation Center, Gainesville, Florida*
[4]*Department of Psychology, University of Florida, Gainesville, Florida*
[5]*Department of Radiology, Memorial Sloan Kettering Cancer Center, New York*
[6]*Department of Rehabilitation Medicine, Atlanta VA Medical Center, Decatur, Georgia*
[7]*Emory University School of Medicine, Atlanta, Georgia*

◆ ══════════════════ ◆

Abstract: Task-correlated motion artifacts that occur during functional magnetic resonance imaging can be mistaken for brain activity. In this work, a new selective detrending method for reduction of artifacts associated with task-correlated motion (TCM) during speech in event-related functional magnetic resonance imaging is introduced and demonstrated in an overt word generation paradigm. The performance of this new method is compared with that of three existing methods for reducing artifacts because of TCM: (1) motion parameter regression, (2) ignoring images during speech, and (3) detrending time course datasets of signal components related to TCM (deduced from artifact corrupted voxels). The selective detrending method outperforms the other three methods in reducing TCM artifacts and in retaining blood oxygenation level dependent signal. *Hum Brain Mapp 30:1105–1119, 2009.* © 2008 Wiley-Liss, Inc.

Key words: FMRI; task-correlated motion; artifact reduction; speech; word-generation; image processing

◆ ══════════════════ ◆

INTRODUCTION

Blood oxygenation level dependent (BOLD) functional magnetic resonance imaging (FMRI) signal changes are of the order of 1–10% of the baseline signal intensity in conventional 1.5T and 3T MR systems. As a result, signal changes from sources (e.g., motion artifacts) other than functional activation must be eliminated as much as possible. Motion can impair the detection of real functional activation in several ways. Random motions of the head appear as additive noise in the voxel time-series and decrease the detectability of functional activation. Task-cor-

related motion (TCM) can mask real activation in cortical regions and can also appear as false positive activation at the brain edges and high-contrast regions of the brain. Proper image registration, most often achieved with intensity-based methods [Cox and Jesmanowicz, 1999; Friston et al., 1995; Woods et al., 1998], is imperative to minimize the effect of motion on the FMRI signal. However, motion artifacts would persist in the FMRI time-series even after the image registration step. One way this occurs is through spin-history effects, where through-plane motion modulates the steady-state magnetization set up by the train of RF excitation pulses in the imaging sequence, causing signal intensity fluctuations [Bullmore et al., 1999; Friston et al., 1996]. Motion can also cause time-dependent modulations in the background magnetic field [Jezzard and Clare, 1999], resulting in artifacts in FMRI time-series images [Haacke et al., 1999; Jezzard and Clare, 1999; Noll, 1991]. Apart from the sources mentioned above, local and/or nonrigid motions can also lead to significant signal changes in the FMRI time-series. These types of signal changes are not linearly related to global rigid motion parameters. In fact, significant signal changes in the FMRI time-series have been observed in a restrained motionless MRI phantom due to the motion of a separate phantom outside the field of view [Yetkin et al., 1996].

FMRI paradigms that require overt word generation [Barch et al., 1999; Crosson et al., 2005; Huang et al., 2002] suffer from task-correlated signal changes arising from speech. Speech involves local nonrigid movement of pharynx, tongue, and jaw, which have been demonstrated to change the magnetic field distribution within the brain [Birn et al., 1998]. Speaking can induce bulk magnetic susceptibility variations due to changes in the volume of oxygen in the airways. Modulation of breathing during speech can also induce task-correlated changes in FMRI signal [Mehta et al., 2006]. The greatest magnetic field changes due to speaking occur in the inferior, temporal, and frontal regions of the brain, decreasing rapidly superiorly and posteriorly. Voxels in the inferior, temporal, and frontal regions may exhibit more than 100% task-correlated signal changes [Kemeny et al., 2005]. After statistical signal processing, these regions often show false positive activation with statistical significance similar to or higher than voxels exhibiting BOLD signal change. FMRI studies involving language have been severely limited by these to speech-related artifacts. Viable techniques to mitigate this artifact will lead to the ability to perform a wider range of language FMRI studies.

Motion-related artifacts in FMRI time-series have been treated retrospectively by a number of different data analysis strategies. One method [Bullmore et al., 1999; Friston et al., 1996] is to model the motion-related signal changes in the FMRI time-series as a function of motion parameter estimates from the image registration program. With this approach the FMRI signal analysis problem is reduced to a multiple linear regression fit of the FMRI time-series data to the modeled BOLD signal changes and the modeled motion-related signal changes. One drawback of motion parameter regression (MPR) and motion parameter examination methods, especially as applied to FMRI of overt speech [Basho et al., 2007; Heim et al., 2006; Palmer et al., 2001], is that there may not be a linear relation between global rigid motion parameters and local signal changes. The motions occurring during speech are nonrigid and local in nature and cause spatially inhomogeneous magnetic field changes.

The major portion of signal changes due to speech-related TCM usually occurs in the first 4–5 s after speaking begins [Birn et al. 1998, 1999, 2004; Mehta et al., 2006]. The BOLD hemodynamic response to cognitive processes typically exhibits a delay of 1–6 s before onset, and the peak is reached 4–6 s after onset [Cohen, 1997; Hoge and Pike, 2001]. This inherent difference in the time scales of BOLD and TCM signal changes has been used to mitigate the effects of TCM due to overt word production by ignoring the first few images during and after speech, either retrospectively in data analysis [Barch et al., 1999; Birn et al., 2004] or prospectively during acquisition [Abrahams et al., 2003; Edmister et al., 1999; Gracco et al., 2005; Hall et al., 1999]. Some related approaches are to screen images (voxels) in which motion is indicated by temporal changes in the signal phase [Huang et al., 2002; Soltysik and Hyde, 2006] or to correct points in the individual voxel time-series that exhibit abnormal deviations [Huang et al., 2008] or to model the TCM signal changes with spike-shaped ideal regressors [Birn et al., 2004].

However, some TCM signal changes can last 8 s or more [Gopinath, 2003; Gopinath et al., 2003; Palmer et al., 2001]. Ignoring the images acquired during speech production does not work well if there is a significant overlap between the time-courses of TCM and hemodynamic response [Birn et al., 1999; Gopinath, 2003; Gopinath et al., 2003], because reduction of TCM artifact compromises the detection and estimation of the BOLD hemodynamic response (HDR). Some brain areas involved in word generation, such as supplementary motor area [Crosson et al. 2003, 2005; Peck et al., 2004], can exhibit BOLD HDRs which start before the stimulus cue [Abdullaev and Posner, 1998; Kraut et al., 2003]. Voxels in these regions are particularly susceptible to overlap of HDR and TCM. Also, in FMRI studies involving patients with language deficits, the time lag between stimulus delivery and overt production can be much longer and more variable than in controls, and the vocalization itself may be protracted [Crosson et al., 2007]. In such cases, ignoring (or screening) images during and after speech can lead to further compromises in detection and estimation of HDR, especially in brain regions where the cognitive processes start before overt production does. Inadequate resolution of the TCM and hemodynamic signal changes also compromises the nonselective global detrending procedure method proposed by Birn et al. [1999], wherein TCM artifact reduction is achieved by orthogonalizing all the voxel time-series in the FMRI dataset with TCM signal changes derived from false positive

voxels near the brain edge. When TCM and HDR signal changes are not temporally well separated, application of this method results in loss of sensitivity for detecting brain activation [Gopinath, 2003; Gopinath et al., 2003].

Some authors [Barch et al., 1999; Bullmore et al., 1999] have argued that the TCM artifacts vary across individuals and thus can be mitigated by averaging across participants. However, speech-related artifacts can arise from vocal tract movements that are generically similar across participants, and hence not reduced by averaging. A recent study found artifactual task-correlated activations during speech in a group of six subjects [Kemeny et al., 2005]. In studies conducted in our lab, it has been observed (unpublished data) that TCM artifacts persist in group-averaged activation maps even up to group sizes of 20. Further, averaging across participants is not applicable in subject-specific case studies such as monitoring brain activation of stroke patients, where variations in lesion topography and the need to visualize activity in perilesional structures prevent group averaging.

In this article, a novel improvement on the TCM detrending method [Birn et al., 1999; Gopinath et al., 2001] is introduced and tested in an event-related overt word generation FMRI paradigm. The difference in the evolution of the speech-related TCM artifact and HDR time courses is used to maximize the retention of BOLD signal and the reduction of TCM-correlated signal changes in the FMRI time-series. Results from this selective detrending method are compared to those obtained from three other methods for reducing TCM: (1) MPR [Bullmore et al., 1999; Friston et al., 1996], (2) ignoring images during speech [Barch et al., 1999] and (3) nonselective detrending [Birn et al., 1999; Gopinath et al., 2001]. These comparisons are made on eight datasets: four obtained from scanning two aphasia patients before and after rehabilitation therapy, and four from scanning four normal control subjects once each, using an event-related overt word generation paradigm. A procedure to compare the four methods in terms of their capacity to reduce TCM artifacts and to retain HDR signal is described. Finally, the implications of this new method for analysis of event-related FMRI data from paradigms susceptible to TCM are discussed.

METHODS

Subjects and Task

The artifact reduction methods were applied on a total of 8 FMRI datasets. These datasets were obtained from two related FMRI studies. Two aphasia patients (one female aged 48 years and one male aged 46 years) with left hemisphere stroke were scanned twice, once before rehabilitation therapy and once after therapy. In the scanner, patients performed an event-related overt word-generation task, for which they were asked to provide single-word responses to semantic category cues. Pseudo-random interstimulus intervals (ISIs) between category cues were

24.9, 26.6, 28.2, or 29.8 s, corresponding to 15, 16, 17, or 18 images, respectively. Four healthy control subjects (three female and one male: ages 50–70 years; mean 57.8 years) were also scanned with a similar event-related overt language FMRI paradigm in which the subjects were asked to provide single-word responses to semantic category cues. Pseudo-random ISIs between category cues was 16.6, 18.3, 19.9, or 21.6 s, corresponding to 10, 11, 12, or 13 images. The ISIs for the patients was longer than those for the control subjects because the aphasics exhibited longer response latencies. Further details about the tasks and parameters are available elsewhere [Crosson et al., 2007; Crosson et al., 2005].

Image Acquisition

MR imaging was performed on a 3T General Electric LX scanner (GE Medical Systems, Waukesha, WI). Low-resolution functional MRIs were obtained using a 1-shot forward spiral sequence [Noll et al., 1995]. Thirty-two 4.0–4.5 mm thick sagittal slices covering the whole brain were acquired. The repetition time (TR) was 1660 ms, the echo time (TE) was 18 ms, the flip angle (FA) was 70°, and the field of view (FOV) was 200 mm. The image matrix size after interpolation was 64×64 and spatial resolution was 3.1 mm $\times$ 3.1 mm $\times$ 4.0–4.5 mm. Five functional runs were acquired with 161 (for patients) or 111 (for controls) images in each run, 1.66 s/image. A time-of-flight MR angiogram (17 ms TR, 4.9 ms TE, 50° flip angle, 256×128 matrix, 200 mm FOV; 4.0–4.5 mm thickness) with the same slice prescriptions as the spiral sequence was acquired for angiographic reference. For anatomic reference, a high resolution T_1-weighted 3D spoiled gradient recalled (SPGR) sequence was acquired with scanning parameters: 240 mm FOV, 256×256 image matrix, 0.9 mm $\times$ 0.9 mm $\times$ 1.3 mm resolution, TR = 23 ms, TE = 7.7 ms and FA = 25°. Auditory stimuli were presented and subject responses were recorded with an MRI-compatible headset and microphone (Resonance Technology, Northridge, CA). Subject responses were coded with Cool Edit™ software (now Adobe® Audition® 2.0, San Jose, CA) into two categories, "correct" and "other" responses. The data analysis was performed with AFNI [Cox, 1996] and Matlab (Mathworks, Sherborn, MA) software.

Data Analysis

The five functional runs were spatially registered to the last image of the last run (closest in time to the angiogram and the T_1-weighted high resolution anatomic) with an intensity based iterated linear weighted least squares algorithm [Cox and Jesmanowicz, 1999] provided in the AFNI software package. The five functional runs were further detrended of linear drifts and concatenated. Since the task paradigm elicited unpredictable and often incorrect responses from the patients, response-locked data analysis [Maccotta et al., 2001] was employed rather than conven-

tional stimulus-locked analysis. To estimate the TCM-related signal changes, all the subject responses (both "correct" and "other") were pooled together and an overt response-locked analysis vector (OAV) was constructed. "Other" responses included "incorrect" responses and unintelligible utterances. For each voxel, the signal was considered to be the result of a convolution of the OAV with the voxel impulse response function (IRF) plus a constant baseline and a white-noise term. IRFs were deconvolved for each voxel from the knowledge of the observed voxel FMRI intensity time-series and the stimulus vector. The significance of activation at each voxel was assessed through the use of a General Linear Model. The significance of activation was assessed through the calculation of the F-statistic for regression. The coefficient of determination R^2 was also calculated.

Characterization of TCM and BOLD Impulse Responses

Representative epochal TCM and BOLD signal changes were selected from the deconvolved voxel IRFs based on prior empirical observations of BOLD MR signal and TCM-related signal. TCM-related signal changes arising from speech exhibit large (sometimes >100%) and sudden signal fluctuations [Birn et al., 1998, 1999; Gopinath et al., 2001; Yetkin et al., 1996]. The amplitude and extent of the TCM artifact is observed to be largest in the frontal, inferior, and temporal regions of and outside the brain and is progressively attenuated towards the posterior, superior, and medial regions. BOLD HDRs are characterized by smaller amplitude (up to ∼10%) and slower changes in signal lasting around 10–20 s after the stimulus event and have been well characterized in literature [Cohen, 1997; Friston et al., 1994; Hoge and Pike, 2001].

Distinct voxel IRFs which were unambiguously TCM-related (from a priori knowledge) at or near high contrast boundaries, on the edge of or outside the brain, that possessed R^2-values above the threshold for inferring brain activation were chosen to characterize TCM IRFs in the FMRI dataset. Appendix A lists the criteria used as well as steps followed to pick the representative TCM IRFs. Figure 1 illustrates this selection procedure for the case of a few TCM IRFs. One lateral (above) and one medial (below) sagittal slice of the spiral FMRI dataset from P1 are shown overlaid with their corresponding R^2-activation maps. The three green arrows point from clusters of TCM-corrupted voxels to screen-grabs of corresponding 2-voxel × 2-voxel grids of IRF curves obtained through AFNI's "Graph" menu. One representative TCM IRF was picked from each of the three clusters illustrated in the figure. This process was repeated for all other slices. Around 10–15 (depending on the dataset) of such distinct TCM IRFs were needed to adequately represent TCM signal changes for the whole-brain FMRI datasets. The chosen TCM IRFs were convolved with the known OAV to form the corresponding TCM vectors. Figure 2a shows some putative TCM-cor-rupted impulse responses chosen from voxels outside the brain for a representative FMRI dataset. Some of the TCM IRFs (e.g. the unmarked dashed curve) are delayed and some (unmarked solid curve) have significantly longer lasting effects. Three to five representative BOLD HDRs spanning all observed HDR delays to onset in the FMRI dataset were also selected. Appendix A lists the criteria used as well as steps followed to pick the representative BOLD HDRs. The two blue arrows in Figure 1 point from BOLD voxels to screen-grabs of corresponding 1-voxel IRF curves obtained through AFNI's "Graph" menu. Figure 2b shows the set of BOLD HDRs chosen from the same dataset.

Selective Detrending Algorithm

From each of the 10–15 representative TCM IRFs, a 805-image (for patients) or a 555-image (for controls) artifact vector was constructed by convolution with the known OAV. For each voxel IRF in the FMRI dataset, the cross-correlation coefficient of the voxel IRF with each of the 10–15 representative TCM IRFs, as well as the cross-correlation coefficient with each of the 3–5 representative BOLD HDRs was calculated. The maximum absolute value (denoted CCT) among the voxel IRF's cross-correlation coefficients with the 10–15 representative TCM IRFs was employed as an indicator of TCM corruption in the voxel time-series. High values of CCT indicated significant TCM corruption. Also, the maximum (denoted CCB) of the voxel IRF's cross-correlation coefficients with the 3–5 representative BOLD HDRs was employed as an indicator of BOLD signal in the voxel time-series. High values of CCB indicated significant BOLD signal in the voxel time-series.

Prior empirical observations of the shapes of BOLD HDRs [Cohen, 1997; Friston et al., 1994; Hoge and Pike, 2001] and epochal TCM-related signal changes [Birn et al., 1999, 2004; Gopinath et al., 2001] were used to formulate criteria for the selective detrending algorithm. These discrimination criteria were chosen to maximize the reduction of TCM-correlated signal as well as to maximize the retention of BOLD signal. Voxels were classified to possess significant TCM-related signal changes, negligible TCM corruption or both BOLD and TCM signal based on the values of CCT and CCB, and treated accordingly. For a given voxel:

$$\text{if } CCT > 0.5 \quad \text{and} \quad CCT - CCB > \tau, \qquad (1)$$

where τ is a separability threshold (a tuning parameter which encodes the tradeoff between reduction of TCM-related signal and retention of BOLD signal), the corresponding voxel time-series was deemed to possess significant and/or predominantly TCM-related signal. Consequently, the voxel time-series was detrended of components proportional to the maximally correlated TCM vector using linear least squares fit (Matlab; Mathworks, Sherborn, MA). Essentially, the linear regression fit to the voxel time-series of the maximally correlated TCM vector was subtracted from the voxel time-series. Voxels not satisfying these crite-
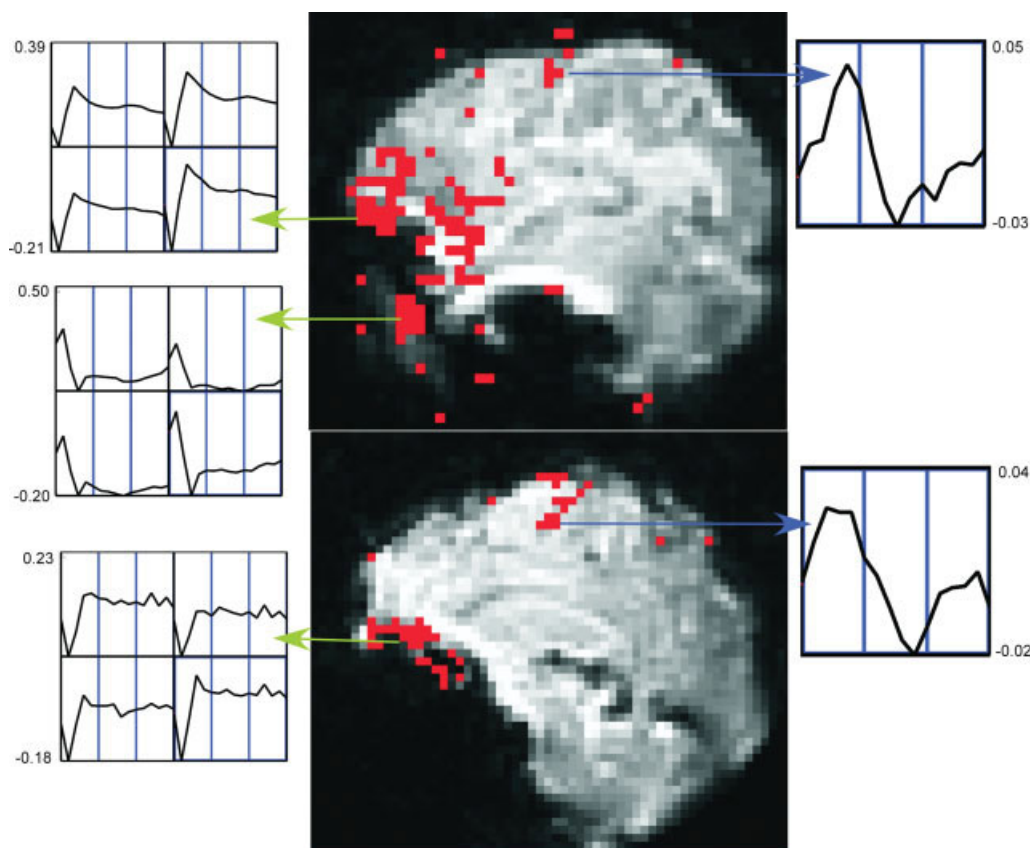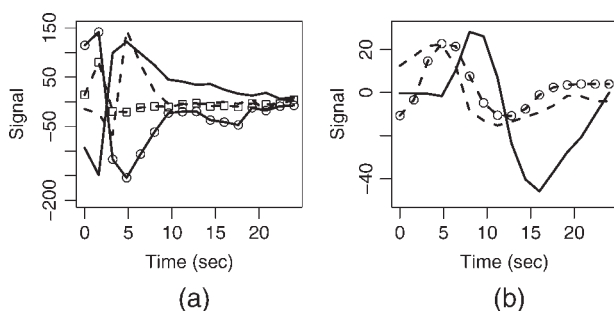
Figure 1.

Estimation of TCM IRFs from a lateral sagittal slice (top) and medial sagittal slice (bottom) for dataset P1. Activation map, thresholded at $R^2 > 0.16$, is overlaid on low-resolution spiral FMR images. The green arrows point from clusters of voxels surmised to be TCM-corrupted and screen-grabs of the corresponding 2×2 voxel grid TCM IRF graphs generated by AFNI.

The blue arrows are drawn between putative BOLD voxels and their corresponding BOLD IRFs. The horizontal axis for each of the individual voxel plots runs from 0 to 24.9 s. The vertical scales for all the individual plots are marked in units of fractional signal change.
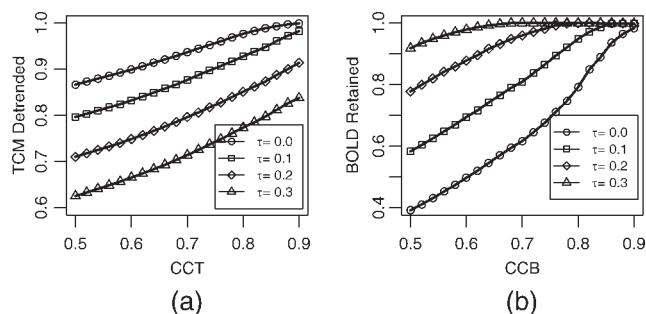
ria were deemed either to possess negligible TCM signal (if CCT < 0.5) or to contain significant amounts of both TCM and BOLD signal, and were left untouched.

Figure 2.

Some representative TCM-related voxel IRFs (**a**); and some representative BOLD HDRs (**b**) chosen from the overt speech FMRI dataset from P1.

The choice of the separability threshold τ in Eq. (1) is governed by a desire to balance the need to detrend as many TCM corrupted voxels (characterized by possessing high values of CCT) as possible while leaving voxels possessing significant BOLD signal (characterized by high values of CCB) intact. Figure 3 elaborates this tradeoff for the dataset P1. Figure 3a shows the fraction of the voxels passing the criteria for detrending in Eq. (1) as a function of TCM corruption indicator CCT, for four different levels of separability threshold τ : 0.0, 0.1, 0.2 and 0.3. Ideally as CCT is increased (indicating strong TCM corruption) the fraction of voxels detrended (passing the criteria in Eq. (1) should approach 1. As the separability threshold τ is increased, the fraction of TCM corrupted voxels (those with high CCT; say CCT > 0.9) that are detrended becomes progressively less. Figure 3b shows the fraction of the voxels that are not detrended (i.e. do not pass the criteria of Eq. (1) as a function of CCB, for the same four levels

Figure 3.
(**a**) Fraction of voxels in dataset P1 detrended as a function of threshold CCT for four different levels of separability, $\tau = 0.0$ (circles), $\tau = 0.1$ (squares), $\tau = 0.2$ (diamonds), and $\tau = 0.3$ (triangles). (**b**) Fraction of voxels in the same dataset left intact as a function of threshold CCB for the same levels of τ (see text in Methods section).

of separability threshold τ (0.0, 0.1, 0.2, and 0.3). Ideally as CCB is increased (indicating strong likelihood of BOLD signal) the fraction of voxels retained should approach 1. Unlike Figure 3a, as the separability threshold τ is increased, the fraction of likely BOLD activation voxels (those with high CCB; say CCB > 0.9) that are retained (not detrended) becomes progressively greater. To balance these opposing constraints, the value of the separability threshold τ, was obtained by finding the value of τ that maximizes the product of fraction of voxels with CCT > 0.8 that are detrended and the fraction of voxels with CCB > 0.7 that are left untouched. The maximization of this product (termed 'selectivity' for the sake of brevity) is illustrated for the same dataset P1 in Figure 4. For all the datasets studied this value of τ remained between 0.17 and 0.23. The choice of threshold CCB > 0.7 and threshold CCT > 0.8 in the process of determining τ weights the selective detrending algorithm towards leaving most voxels with BOLD signal undetrended at the cost of leaving some TCM corrupted voxels undetrended. For instance, if all voxels with CCT > 0.9 were TCM corrupted (which is almost certain) and all voxels with CCB > 0.9 possessed BOLD signal, from Figure 3 this algorithm will leave 15% of the TCM corrupted voxels untouched while detrending less than 5% of voxels which possess BOLD signal.

The detrended time-series were then re-analyzed with the GLM-based deconvolution analysis described above. The activation due to "other" responses was modeled as a nuisance signal and the partial F statistic of the "correct" response related signal and the corresponding R^2-coefficient quantified brain activation.

Operator Dependence of Selective Detrending

Since the selective detrending method requires interactive selection of putative TCM IRFs and BOLD HDRs by the operator, potential operator-dependent variability is a

concern. To test the operator dependence of the selective detrending algorithm, five different individuals carried out the steps in the algorithm on the eight FMRI datasets mentioned above. The operators (except one) were not involved in the development of the algorithm and had varying levels of experience with FMRI data analysis. The independent operators were given instructions on how to pick representative TCM IRFs and BOLD HDRs (as described in Appendix A) and trained on a separate FMRI dataset (not included in the study) and left to analyze the eight datasets without further input. The Pearson product moment correlation coefficient was used to assess reproducibility of the results of selective detrending (artifact-reduced FMRI activation R^2 statistic) across the five operators for each of the eight datasets.

Other Methods

All the other methods were implemented under GLM-based deconvolution analysis framework in the AFNI software package so that the results from various methods could be compared consistently. The motion-parameter regression method [Bullmore et al., 1999; Friston et al., 1996] was implemented by obtaining the motion parameters for each image in the time-series from the image-registration program. For each voxel, the signal at each time-point related to rigid motion was modeled as a quadratic function of the motion parameters in that particular time-point and the preceding time-point. The motion-related fluctuations along with the "other" response related signal changes were modeled as nuisance signal, and the partial F statistic of the "correct" response related signal and the corresponding R^2-coefficient quantified brain activation. The method of screening images during speech [Barch et al., 1999; Birn et al., 2004] was implemented by ignoring the zeroth and first lags while estimating the voxel IRF by deconvolution analysis, i.e. the minimum lag number was set to 2 ('ML2'). The nonselective detrending method [Birn et al., 1999; Gopinath et al., 2001] was implemented in a
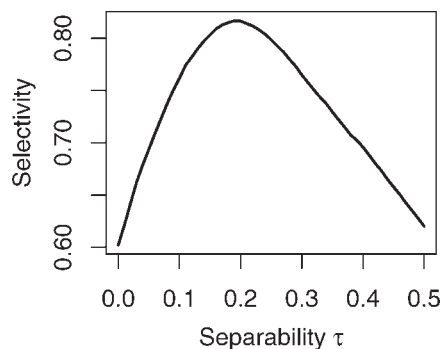
Figure 4.
Maximization of 'selectivity,' the product of the fraction of TCM voxels detrended and BOLD voxels retained as a function of separability τ (see text in Methods section) for dataset, P1.

manner similar to the selective detrending method up to the part regarding estimation of TCM IRFs. In the nonselective detrending method, all the voxels in the FMRI dataset were detrended (using linear least-squares fit) of components proportional to all the chosen representative TCM vectors. The detrended time-series were then reanalyzed with the GLM-based deconvolution analysis similar to the selective detrending method. In this respect this method differs from the earlier published nonselective detrending method of Birn, [Birn et al. 1999] which utilized multiple regression with a fixed shape hemodynamic response model. This variant of the non-selective detrending method has been introduced in abstract form [Gopinath et al. 2001]. The number of TCM vectors used in the nonselective detrending method was determined by what was needed to detrend TCM signal from all supra-threshold TCM corrupted voxels outside the brain. Since quite a few of the chosen TCM vectors used in selective detrending were linear combinations of each other, about 5 or 6 TCM vectors were sufficient for non-selective detrending. Thus, the number of degrees of freedom lost in the detrending is about 5 or 6, which is not significant given the large number of images (805 or 555).

Comparison of TCM Artifact Reduction Methods

Observations on the spatial and temporal characteristics of the TCM artifacts [Barch et al., 1999; Birn et al., 1998,

1999, 2004; Gopinath, 2003; Gopinath et al., 2001; Huang et al., 2002; Huang et al., 2008] and BOLD activation [Cohen, 1997; Friston et al., 1994; Hoge and Pike, 2001] were used to create a pool of putative BOLD voxels and a pool of putative TCM voxels for each dataset. The performance of the four TCM artifact reduction methods was compared in terms of the reduction in significance of "activation" of the putative TCM voxels and retention of the significance of activation of putative BOLD voxels.

AFNI software was used to create intensity-based brain and nonbrain masks for each FMRI dataset. Voxels outside the brain with R^2 or F values above the threshold selected to infer brain activation and exhibiting larger fluctuation in the first five images than the last 11 (8 for controls) were assigned to the pool of putative TCM-voxels. As a final level of screening, voxels satisfying the above criteria and activating above a threshold for R^2 ($R^2 > 0.05$) were examined visually to further ensure absence of BOLD-like voxels in this pool of putative TCM-voxels.

For constructing the pool of putative BOLD voxels, the intra-cerebral voxel IRFs, specified by the brain masking method to ignore voxels outside the brain and near brain edges, were fitted to a generic hemodynamic response curve. This curve, which is included in the documentation of the AFNI software package (under the program *waver*) is parameterized in terms of amplitude, delay-time, rise-time, fall-time, restore-time, and undershoot (as a fraction of amplitude). The functional form of the curve is given by

$$
\begin{aligned}
h(t) &= \frac{A}{2}\left[\tanh\left[\tan\left(\frac{\pi}{2}\left(\frac{1.6(t-t_{\mathrm{D}})}{t_{\mathrm{R}}}-0.8\right)\right)\right]+1\right] && \text{for } t_{\mathrm{D}} < t \le t_{\mathrm{D}}+t_{\mathrm{R}} \\
h(t) &= \frac{A}{2}\left[(1+U)\tanh\left[\tan\left(\frac{\pi}{2}\left(\frac{1.6(t_{\mathrm{F}}+t_{\mathrm{g}}+t_{\mathrm{D}}-t)}{t_{\mathrm{F}}}-0.8\right)\right)\right]-U\right] && \text{for } t_{\mathrm{D}}+t_{\mathrm{R}} < t \le t_{\mathrm{D}}+t_{\mathrm{R}}+t_{\mathrm{F}} \\
h(t) &= \frac{A}{2}\left[(-U)\tanh\left[\tan\left(\frac{\pi}{2}\left(\frac{1.6(t_{\mathrm{F}}+t_{\mathrm{g}}+t_{\mathrm{D}}+t_{\mathrm{RT}}-t)}{t_{\mathrm{RT}}}-0.8\right)\right)\right]\right] && \text{for } t_{\mathrm{D}}+t_{\mathrm{R}}+t_{\mathrm{F}} < t \le t_{\mathrm{D}}+t_{\mathrm{R}}+t_{\mathrm{F}}+t_{\mathrm{RT}} \\
h(t) &= 0 && \text{otherwise}
\end{aligned}
\tag{2}
$$

where A is the amplitude, t_{D}, t_{R}, t_{F} and t_{RT} are delay-time, rise-time, fall-time, and restore-time of the hemodynamic response, respectively, and U is the amplitude of the post-stimulus undershoot as a fraction of A. The advantage of this function over the conventional γ-variate functions commonly employed in FMRI literature [Cohen, 1997; Friston et al., 1994] is that it is parameterized in terms of different phases and temporal aspects of the HDR, which could be used to distinguish between BOLD and TCM voxels, as described below. The IRF of each voxel was fitted to the generic HDR by means of a nonlinear optimization method using the *lsqcurvefit* function of Matlab™.

The maxima of the first four images of each voxel IRF, $\max(\mathrm{IRF}_{1-4})$ and the corresponding minima, $\min(\mathrm{IRF}_{1-4})$ were also calculated. Intra-cerebral voxel IRFs satisfying the following criteria formed the pool of putative BOLD voxels

$$
\begin{aligned}
&A > 0 \\
&t_{\mathrm{D}}+t_{\mathrm{R}}+t_{\mathrm{F}} > 6 \\
&t_{\mathrm{D}}+t_{\mathrm{R}} < 11 \\
&t_{\mathrm{R}}+t_{\mathrm{F}} > 4 \\
&t_{\mathrm{R}} > 1 \\
&t_{\mathrm{F}} > 2 \\
&t_{\mathrm{D}} < 4 \\
&\max(\mathrm{IRF}_{1-4}) > 1.2|\min(\mathrm{IRF}_{1-4})|
\end{aligned}
\tag{3}
$$

where the terms are as defined above. Time is expressed in units of image TR (TR ≈ 1.66 s). This set of criteria was chosen to keep IRFs that closely resemble BOLD HDR shapes reported in the literature [Cohen, 1997; Friston et al., 1994; Hoge and Pike, 2001] while at the same time eliminating TCM-related IRFs from contaminating the pool of putative BOLD voxels. These criteria were devised to

keep voxels with slower signal changes characteristic of BOLD HDRs in the pool while screening out voxels with large deviations in signal change in the initial phases. These phenomenological constraints on the HDR parameters were arrived at after observing the fitted values of these parameters when applying Eq. (2) to the chosen TCM IRFs for all the eight datasets. As a final level of screening, voxels satisfying the criteria in Eq. (3) and activating above a threshold ($R^2 > 0.05$) were visually inspected to ensure absence of TCM-like voxels in the pool of putative BOLD voxels. This method of obtaining test-beds of putative BOLD voxels rejects a number of voxels possessing BOLD responses (e.g. HDRs, whose positive phases are shorter than 10 s full-width at baseline) from the BOLD test-bed. This automated method was also slightly biased against the performance of the selective detrending method, as it ignored a number of voxels in which the selective detrending method outperformed the non-selective detrending method and the method of ignoring images in terms of BOLD signal retention. However, the resultant pool was devoid of TCM-corrupted voxels.

For all the four methods as well as for the case of no motion correction, the distributions of the number of voxels in the putative TCM and BOLD voxel pools detected as a function of threshold R^2 were constructed to provide a means of comparing the capabilities of the four methods to retain putative BOLD voxels and eliminate putative TCM-corrupted voxels. The number of voxels registered "active" (at $R^2 > 0.16$) in the selective detrending method ('dtsel') and the other three methods ('MPR,' 'ML2,' and 'dtr') were tabulated (Tables I and II) as a fraction of the number of voxels found "active" in the untreated dataset ('ML0'). Summary statistics, in the form of paired t-tests between the fractional number of voxels registered "active" (at $R^2 > 0.16$) in the selective detrending method ('dtsel') and the other three methods ('MPR,' 'ML2,' and 'dtr'), were computed for the data in Table I (TCM artifact reduction) and Table II (BOLD voxels retention) across the

TABLE I. Proportion of voxels in the putative TCM voxels pool declared "active" $R^2 > 0.16$ ($P < 10^{-7}$) in each of the 8 overt speech FMRI datasets, after each of the four TCM artifact reduction methods: motion parameter regression, 'MPR,' ignoring two images after speech, 'ML2,' nonselective detrending, 'dtr' and selective detrending, 'dtsel' as a fraction of the number of voxels declared "active" in the untreated case, 'ML0'

Dataset	ML0	MPR	ML2	dtr	dtsel
C1	1	0.6542	0.0446	0.002	0.0071
C2	1	0.4972	0.1959	0.0002	0.0398
C3	1	0.0763	0.0967	0.0053	0.0191
C4	1	0.571	0.2509	0.0005	0.0015
P1	1	0.9379	0.0787	0.0021	0.0083
P2	1	0.1329	0.1798	0.0062	0.0128
P3	1	0.4396	0.2954	0.0008	0.0008
P4	1	0.6842	0.1491	0.0008	0.004

TABLE II. Proportion of voxels in the putative BOLD voxels pool declared "active" $R^2 > 0.16$ ($P < 10^{-7}$) in each of the 8 overt speech FMRI datasets, after each of the four TCM artifact reduction methods: motion parameter regression, 'MPR,' ignoring two images after speech, 'ML2,' nonselective detrending, 'dtr' and selective detrending, 'dtsel' as a fraction of the number of voxels declared "active" in the untreated case, 'ML0'

Dataset	ML0	MPR	ML2	dtr	dtsel
C1	1	0.8969	0.8493	0.0274	0.9589
C2	1	0.9294	0.8565	0.0024	0.9718
C3	1	0.5879	0.8478	0.1386	0.974
C4	1	0.7699	0.8849	0.0885	0.9292
P1	1	0.9578	0.8397	0.0802	0.865
P2	1	0.6923	0.804	0.4169	0.8958
P3	1	0.855	0.8874	0.1947	0.9103
P4	1	1	0.7923	0.1444	0.9225

eight datasets. An important property of the pools was that the methods used to obtain them were independent of the algorithm to estimate artifactual TCM signal changes.

RESULTS

Figure 5 serves both to show the extent of TCM-corruption of an overt speech FMRI dataset and also to support the rationale behind the method of choosing TCM IRFs. Figure 5a shows a sagittal slice (centered 29 mm right of midline) of the spiral FMRI dataset from P2. The overlay is a map of the proportionality constant α, in the model fit: voxel IRF = α × maximally correlated TCM IRF. The absolute value of the proportionality constant α (estimated through least squares linear regression) is displayed and the color-coding in the figure goes from blue ($|\alpha| = 0$) to red ($|\alpha| \geq 1$). All the voxels shown with color overlay exhibited maximal cross-correlation coefficient, CCT > 0.9. The overlay map can be thought of as the ratio of the amplitude of the TCM artifact in the voxel IRF to the amplitude of the maximally correlated TCM IRF. Figure 5b,c are from progressively lateral sagittal slices 37 mm and 49 mm left of midline respectively. From Figure 5 it is apparent that the TCM artifact is strongest in the inferior, lateral, temporal and frontal parts of the image, consistent with the speech related magnetic field disturbances reported in Birn et al., [1998]. The TCM IRFs were picked from high contrast boundaries at the brain edges or outside the brain, but exhibit high correlation with a significant number of voxels inside the brain. The amplitude of the artifact in the TCM corrupted voxels inside the brain is attenuated with respect to the representative TCM IRFs chosen from voxels outside the brain. Figure 6a,b shows some representative TCM IRFs chosen for the datasets C1 and C2, respectively. TCM IRFs showed a good degree of commonality between different datasets. However, some distinct TCM IRFs were also found in some datasets, especially in stroke patients.

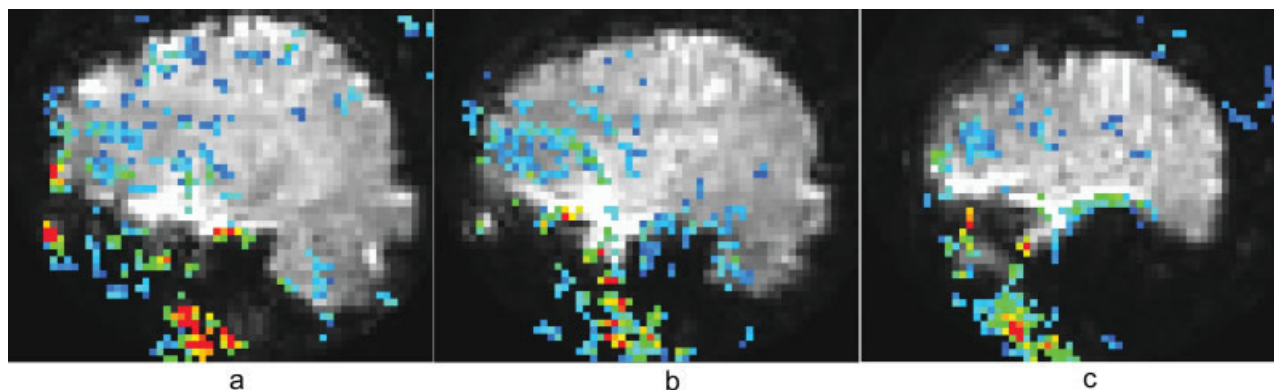Visual inspection of the R^2-activation maps indicates superior performance of the selective detrending method

Figure 5.

Individual sagittal slice maps of the absolute value of the proportionality constant α (see Results section) representing the intensity of the TCM artifact, overlaid on their corresponding slices of the spiral FMRI dataset from P2. All the voxels exhibit maximal cross-correlation coefficient, CCT > 0.9. The color-coding is from blue ($|\alpha| = 0$) to red ($|\alpha| \geq 1$).

when compared to the other methods. Figure 7a shows a lateral sagittal slice in the left hemisphere of a high-resolution $T1$-weighted anatomic image with the R^2-activation map overlay for a representative FMRI dataset (P1), without TCM artifact reduction, thresholded at $R^2 > 0.16$ ($P < 10^{-7}$). Voxels in blue were TCM-corrupted and exhibited IRFs similar to those shown in Figures 2a and 6. Voxels in red possessed BOLD HDRs similar to Figure 2b and highlight activation in the cortex along the inferior frontal sulcus, premotor and primary motor cortex. The BOLD activation patterns were consistent with neuropsychological theory of semantic word generation and are described in another publication [Crosson et al., 2005]. Figures 7b–e show the R^2-activation map for the same dataset after MPR, after ignoring the first two time-points after speech ('ML2'), after nonselective detrending and after selective detrending respectively. In this example, selective detrending (7e) dramatically outperforms the MPR method (7b) in TCM artifact reduction. The MPR method did not have an appreciable effect on the TCM-corrupted voxels (or BOLD voxels) inside the brain. The MPR method did seem to

attenuate TCM-artifact outside the brain (data not shown), where the signal changes due to motion was more linearly related to the estimated motion parameters. The method of ignoring the first two time-points after speech (7c) performs similar to the selective detrending (7e) in terms of retention of BOLD activation but performs slightly worse in reducing TCM artifacts. The nonselective detrending method (7d) removes some BOLD signal along with TCM artifacts, leading to artificial loss of brain activation in some voxels.

Figure 8a shows a lateral sagittal slice in the right hemisphere of a high-resolution $T1$-weighted anatomic image with the R^2-activation map overlay for another representative FMRI dataset (P2), without TCM artifact reduction, thresholded at $R^2 > 0.16$ ($P < 10^{-7}$). Figures 8b–e show the R^2-activation map for the same dataset after MPR, after ignoring the first two time-points after speech ('ML2'), after nonselective detrending and after selective detrending respectively. The voxels in blue were TCM-corrupted and voxels in red reflected brain activation in areas (including the inferior frontal gyrus and anterior temporal lobe as well as premotor and primary motor cortex) consistent with the language FMRI paradigm [Crosson et al., 2005, 2007]. TCM artifacts persist in a number of voxels even after the ignoring images during speech (8c). The MPR method (8a) performs better than the method of ignoring images during speech in terms of TCM artifact reduction but at the price of reducing significance of activation in some BOLD voxels. Selective detrending (8e) performs noticeably better at TCM artifact reduction while at the same time retaining BOLD voxels. Nonselective detrending (8d) results in loss of significance of activation in all BOLD voxels along with all TCM corrupted voxels in this example.

Figure 9 shows the deconvolved BOLD HDRs from a voxel in the supplementary motor area (SMA) of an aphasia patient FMRI dataset. The selective detrending
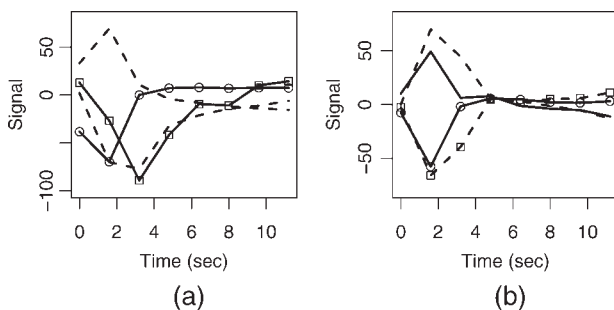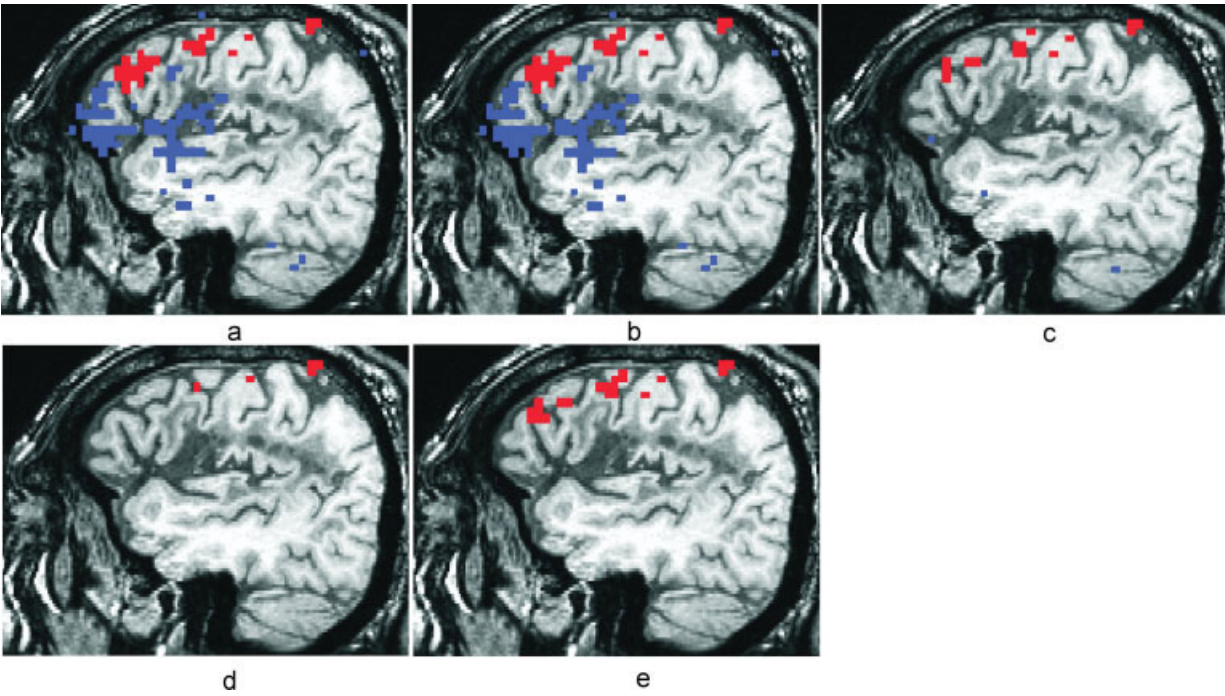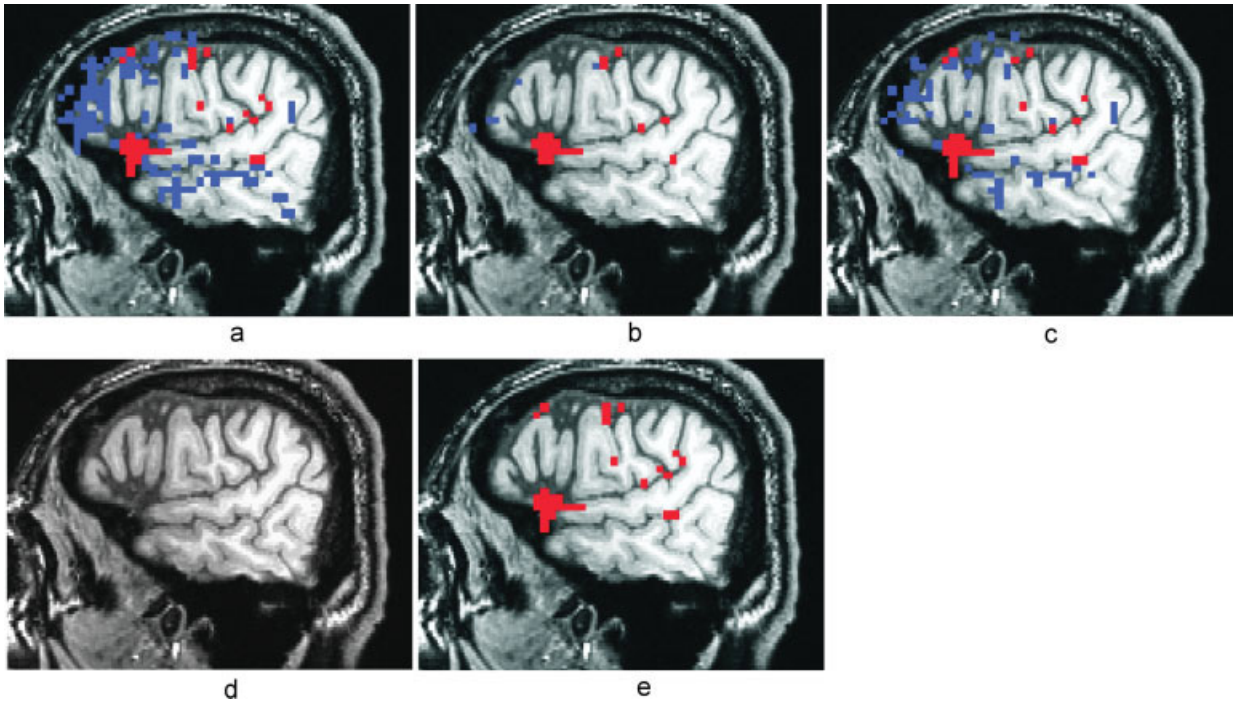

Figure 6.

Some representative TCM-related voxel IRFs from **(a)** dataset C1 and **(b)** dataset C2.

Figure 7.

Activation maps thresholded at $R^2 > 0.16$ ($P < 10^{-7}$) for a representative overt speech FMRI dataset (P1) with no TCM artifact-reduction (**a**), with motion parameter regression (**b**), with ignoring two images after speech (**c**), with nonselective detrending (**d**) and with selective detrending (**e**). Voxels in blue exhibited characteristic TCM IRFs and voxels in red exhibited characteristic BOLD HDRs. All the maps are from the same lateral sagittal slice in the right hemisphere. BOLD areas of activity include cortex along the inferior frontal sulcus, premotor, and primary motor cortex.

Figure 8.

Activation maps thresholded at $R^2 > 0.16$ ($P < 10^{-7}$) for a representative overt speech FMRI dataset from an aphasic subject (P2) with no TCM artifact-reduction (**a**), with motion parameter regression (**b**), with ignoring two images after speech (**c**), with nonselective detrending (*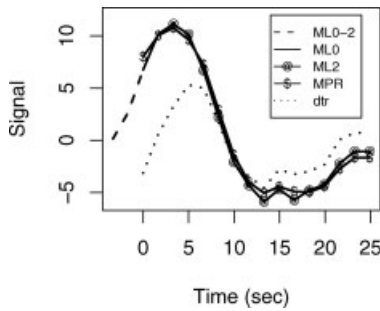*d**) and with selective detrending (**e**). Voxels in blue exhibited characteristic TCM IRFs and voxels in red exhibited characteristic BOLD HDRs. All the maps are from the same lateral sagittal slice in the left hemisphere. BOLD areas of activity include the inferior frontal gyrus and anterior temporal lobe as well as premotor and primary motor cortex.

Figure 9.

Deconvolved BOLD HDRs from a voxel in the supplementary motor area (SMA) of an aphasia patient without TCM artifact reduction, 'ML0' (blue), after motion parameter regression, 'MPR' (green), after ignoring two images after speech, 'ML2' (magenta) and after nonselective detrending, 'dtr' (red). The selective detrending algorithm left this voxel untreated, with the same HDR as 'ML0'. The blue dashed curve ('ML0-2') shows the extension of the BOLD HDR up to 2 image TRs before vocal response enunciation.

algorithm left signal in this voxel intact. The BOLD HDR (blue curve; 'ML0-2') starts before the word generation in this voxel, as shown by the extension of the HDR (dashed blue curve) before the overt response onset (at $t = 0$). This causes the 'ML2' method of ignoring two images after speech (magenta curve) to miss substantial parts of the HDR. Non-selective detrending 'dtr' (red curve) results in attenuation of BOLD signal. The MPR method (green curve) has a negligible effect on the HDR in this voxel.

Tables I and II provide a more comprehensive comparison of the performance of the four methods of TCM artifact reduction. Table I shows the proportion of voxels in the pool of TCM-corrupted voxels with $R^2 > 0.16$ ($P < 10^{-7}$), the threshold selected for inferring brain activation for the four methods of artifact reduction. The proportion of voxels for each method for each dataset is expressed as a fraction of the number of voxels with $R^2 > 0.16$ for the case with no TCM artifact reduction ('ML0'). Table II shows the results of a similar analysis for the pool of BOLD voxels in each of the 8 datasets. As seen from the results in Table I, the selective detrending method ('dtsel') performed substantially better than both the method of ignoring the first two images after vocal response ('ML2') and the MPR method (MPR) in all the datasets (paired t-test: $t_7 < -4.7$, $P < 0.002$). The selective detrending method ('dtsel') performed almost as well as the nonselective detrending method ('dtr') in all the datasets (paired t-test: $t_7 \sim 2.1$, $P < 0.08$). However, as seen in Table II, the selective detrending method was much superior to the nonselective method in retaining putative BOLD activation, retaining significantly more voxels in the BOLD pool (paired t-test: $t_7 > 15$, $P < 10^{-6}$). Table II also shows that the selective detrending method performs better than the method of ignoring images during speech ('ML2') in terms

of retention of BOLD voxels (paired t-test: $t_7 > 3$, $P < 0.01$) and similar to the MPR method (paired t-test: $t_7 \sim 1.7$, $P < 0.13$). The relative performance of the MPR method and the 'ML2' method seems to vary across datasets. For instance the 'ML2' method performs much better at TCM artifact reduction than the MPR method in the dataset P1, but MPR outperforms 'ML2' in the dataset P2. This can also be observed in Figures 7 and 8.

Since the selective detrending procedure involves operator-dependent selection of putative TCM IRFs and BOLD HDRs, it is important to examine the degree of variability in the results obtained by different operators. Table III summarizes the Pearson moment cross-correlation coefficients between the results of the final artifact-reduced R^2-maps obtained by each operator with all the other four operators on the datasets with most (unshaded) and least (shaded) inter-operator correspondence. All the raters exhibited a very high degree of correlation with results obtained by other raters for all the datasets (Pearson moment cross-correlation $P \sim 0$). There is very close correspondence between the selective detrending results obtained by different operators.

DISCUSSION

The extent of TCM-corruption of an overt speech FMRI dataset is apparent from Figure 5. A number of voxels inside the brain exhibit impulse responses very similar to characteristic TCM IRFs selected from voxels at or near high contrast boundaries outside the brain. Some of the selected TCM IRFs are illustrated in Figures 2a and 6a,b. Incidentally, TCM IRFs found inside the brain exhibited a scaled down version of the behavior illustrated in Figure 2a and 6a,b. Thus methods proposed in literature [Huang et al., 2008] which screen for TCM artifacts by detecting inordinate spikes in the voxel time-series may be unable to alleviate some TCM artifacts inside the brain. TCM IRF shapes exhibited a good degree of commonality within and across FMRI datasets as can be seen from Figure 2a and 6a,b. However, distinct TCM IRFs were also occasion-

TABLE III. Pearson product moment correlation coefficient between the selective detrended R^2-maps obtained by five different operators, O1, O2, O3, O4, and O5, with each of the other operators on the 8 FMRI datasets: results for datasets with most (unshaded) and least (shaded) inter-operator correspondence

Operators	O1	O2	O3	O4	O5
O1		0.9394	0.9506	0.8829	0.9168
O2	0.9649		0.9406	0.8993	0.8979
O3	0.9641	0.9723		0.8901	0.9074
O4	0.9293	0.9163	0.9244		0.8781
O5	0.9817	0.9876	0.971	0.9204	

All p-vals ~ 0.

ally found, especially in data of patients with large ische-mic stroke lesions.

Overall, the MPR method was the poorest method of the four tested for reducing TCM artifact. This indicates that the rigid motion parameters do not adequately encode the motion-related changes in FMRI signal during speech. For datasets in which the MPR method performed relatively well in reducing TCM artifact, it performed poorly in retaining BOLD signal. For example, for the dataset C3, the MPR method adequately reduces TCM-related artifact in 93% of the voxels in the putative TCM artifacts pool. However, in dong so, it reduces BOLD activation to below threshold levels in 40% of the putative BOLD voxels. In contrast, for the dataset P1, where the MPR method per-forms well in retaining BOLD signal, leaving intact 95% of the voxels in the putative BOLD voxels pool, it performs poorly in reducing TCM artifact, leaving 93% of the puta-tive TCM voxels artifact-ridden. The performance charac-teristics of the MPR method seem to depend on the degree to which the global motion is related to the time-series of vocal response events. For instance, datasets that exhibit MPR method performance characteristics similar to C3 also display a significant linear relationship between the inferior-superior (I-S) and pitch motion parameter vectors and the OAV. The global I-S and pitch motion vectors of C3 were strongly related to C3s OAV (linear regression $P < 10^{-10}$). On the other hand for P1, which does not respond significantly to MPR treatment, the global I-S and pitch motion vectors were only weakly related to the OAV (linear regression $P > 0.1$). Thus the MPR method seems to have a significant impact on the data when there is a strong linear relationship between the rigid global motion of the head and the task performance, in which cases, the effect of the MPR method is to reduce both the TCM-related artifactual signal as well as BOLD signal due to brain activation.

The method of ignoring images during speech ('ML2') performed better than the MPR method in reducing TCM artifact (paired t-test $P < 0.02$) but underperformed both the detrending methods. The ML2 artifact reduction method relies on the changes in FMRI signal occurring in the first two images during after vocalization. As apparent from Figures 2a and 6a,b some of the TCM IRFs are delayed and/or prolonged in duration. The delay to onset is often more than an image (1.66 s, in our case). This delay may be caused by inter-epoch variance in the TCM artifact, since it depends in part on enunciation [Birn et al., 1999], which could vary from epoch to epoch. The pro-longed TCM artifacts may be due to changes in the vocal apparatus that have a slower time course, such as muscle relaxation [Gopinath, 2003; Gopinath et al., 2003; Palmer et al., 2001]. The ML2 method fails to adequately attenuate the TCM-related artifacts in these instances. As apparent from Figure 9, the ML2 method also masks the BOLD sig-nal from brain activation in areas where the BOLD HDR onsets prior to enunciation. Thus, on the whole the selec-tive detrending method is preferable.

The nonselective detrending method ('dtr') performs bet-ter than all other methods in TCM artifact reduction. How-ever, it performs poorly in retaining BOLD activation. Fig-ures 2 and 9 point towards the source of the suboptimal of BOLD signal retention of this method. Since the selected TCM IRFs span a appreciable range of delays and dura-tions, the necessary condition of separability of the brain activation-related BOLD HDRs from of all chosen TCM IRFs is violated, leading to unsatisfactory performance.

The effect of selective detrending on the Type I error rate is expected to be minimal. Although different voxels in a given dataset are not detrended with the same vector, the process of detrending, which entails loss of only one degree of freedom, is not expected to affect either the acti-vation statistic's voxel-wise univariate or overall mass uni-variate null distributions significantly due to the large number of time-points in the analysis: 805 for patients and 555 for controls. Type II error estimation that relies on the "active" distribution also should not be significantly affected due to the loss of only one degree of freedom in the voxels undergoing selective detrending. The process of manually selecting the representative TCM and BOLD responses takes an average of thirty minutes per dataset. This process benefits from utilization of neuroimaging soft-ware with a good graphical interface (e.g. AFNI), where the operator has easy visual access to voxel-level time-se-ries data. The differences in spatial and temporal origin and signature between TCM IRFs and BOLD HDRs, and the degree of commonality in the shape of TCM IRFs both within and across datasets, indicates that there may be potential for automation in the process of selecting repre-sentative TCM and BOLD responses. The Matlab selective detrending program runs in about 20 min CPU time on a $64 \times 64 \times 32 \times 805$ time-series dataset on a 2 GHz dual processor Linux workstation. From our experience the added processing time is a penalty worth incurring, given the superior performance of the selective detrending method.

CONCLUSION

A new selective detrending method for reduction of arti-fact associated with TCM in event-related FMRI paradigms using an overt word generation task has been introduced and tested. It performs uniformly better than three other commonly used TCM artifact-reduction methods in retain-ing BOLD signal and reducing TCM artifacts. The data suggest that the selective detrending algorithm can be automated for more convenient implementation and use in the future.

ACKNOWLEDGMENTS

REFERENCES

Abdullaev YG, Posner MI (1998): Event-related brain potential imaging of semantic encoding during processing single words. Neuroimage 7:1–13.

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

Barch DM, Sabb FW, Carter CS, Braver TS, Noll DC, Cohen JD (1999): Overt verbal responding during fMRI scanning: Empirical investigations of problems and potential solutions. Neuroimage 10:642–657.

Basho S, Palmer ED, Rubio MA, Wulfeck B, Muller RA (2007): Effects of generation mode in fMRI adaptations of semantic fluency: Paced production and overt speech. Neuropsychologia 45:1697–1706.

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

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

Birn RM, Cox RW, Bandettini PA (2004): Experimental designs and processing strategies for fMRI studies involving overt verbal responses. Neuroimage 23:1046–1058.

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

Cohen MS (1997): Parametric analysis of fMRI data using linear systems methods. Neuroimage 6:93–103.

Cox RW (1996): AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

Cox RW, Jesmanowicz A (1999): Real-time 3D image registration for functional MRI. Magn Reson Med 42:1014–1018.

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

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 (2005): Role of the right and left hemispheres in recovery of function during treatment of intention in aphasia. J Cogn Neurosci 17:392–406.

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

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

Friston KJ, Jezzard P, Turner R (1994): Analysis of functional MRI time-series. Hum Brain Mapp 1:153–171.

Friston KJ, Ashburner J, Frith CD, Poline J-B, Heather JD, Frackowiak RS (1995): Spatial registration and normalization of images. Human Brain Mapp 3:165–189.

Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R (1996): Movement-related effects in fMRI time-series. Magn Reson Med 35:346–355.

Gopinath KS (2003): Reduction of noise due to task correlated motion in event related overt word generation functional magnetic resonance imaging paradigms [PhD thesis]. Gainesville: University of Florida.

Gopinath KS, Briggs RW, Soltysik D, Himes N, Crosson BA (2001): Reduction of noise associated with stimulus correlated motion in event related overt word production fMRI studies. Proc Intl Soc Magn Reson Med 9:297.

Gopinath KS, Peck KK, Soltysik DA, Crosson BA, Briggs RW (2003): 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 Reson Med 11:388.

Gracco VL, Tremblay P, Pike B (2005): Imaging speech production using fMRI. Neuroimage 26:294–301.

Haacke EM, Brown RW, Venkatesan R (1999): Magnetic Resonance Imaging: Physical Principles and Sequence Design. New York: Wiley-Liss. 914 p.

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

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

Hoge RD, Pike GB (2001): Quantitative measurement using functional MRI. In: Jezzard P, Matthews PM, Smith SM, editors. Functional MRI: An Introduction to Methods, 1st ed. Oxford: Oxford Univeristy Press. p 408.

Huang J, Carr TH, Cao Y (2002): Comparing cortical activations for silent and overt speech using event-related fMRI. Hum Brain Mapp 15:39–53.

Huang J, Francis AP, Carr TH (2008): 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 104:10–23.

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

Kemeny S, Ye FQ, Birn R, Braun AR (2005): Comparison of continuous overt speech fMRI using BOLD and arterial spin labeling. Hum Brain Mapp 24:173–183.

Kraut MA, Calhoun V, Pitcock JA, Cusick C, Hart J Jr (2003): Neural hybrid model of semantic object memory: Implications from event-related timing using fMRI. J Int Neuropsychol Soc 9:1031–1040.

Maccotta L, Zacks JM, Buckner RL (2001): Rapid self-paced event-related functional MRI: Feasibility and implications of stimulus-versus response-locked timing. Neuroimage 14:1105–1121.

Mehta S, Grabowski TJ, Razavi M, Eaton B, Bolinger L (2006): Analysis of speech-related variance in rapid event-related fMRI using a time-aware acquisition system. Neuroimage 29:1278–1293.

Noll DC (1991): Reconstruction techniques for magnetic imaging [Phd Thesis]. Palo Alto: Stanford University.

Noll DC, Cohen JD, Meyer CH, Schneider W (1995): Spiral K-space MR imaging of cortical activation. J Magn Reson Imaging 5:49–56.

Palmer ED, Rosen HJ, Ojemann JG, Buckner RL, Kelley WM, Petersen SE (2001): An event-related fMRI study of overt and covert word stem completion. Neuroimage 14(1 Part 1):182–193.

Peck KK, Moore AB, Crosson BA, Gaiefsky M, Gopinath KS, White K, Briggs RW (2004): Functional magnetic resonance imaging before and after aphasia therapy: Shifts in hemodynamic time to peak during an overt language task. Stroke 35:554–559.

Soltysik DA, Hyde JS (2006): Strategies for block-design fMRI experiments during task-related motion of structures of the oral cavity. Neuroimage 29:1260–1271.

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

Yetkin FZ, Haughton VM, Cox RW, Hyde J, Birn RM, Wong EC, Prost R (1996): Effect of motion outside the field of view on functional MR. AJNR Am J Neuroradiol 17:1005–1009.

APPENDIX A

Operator Instructions for Picking Representative TCM IRFs and Representative Bold HDRs

1. Keep the 805 (555 for controls) image 3D+time spiral FMRI time-series as the underlay file and the R^2 statistic file (thresholded at $R^2 > 0.16$ for datasets P1–P4 and $R^2 > 0.20$ for datasets C1–C4) as overlay.

2. Examine all supra-threshold voxels outside the brain or at high contrast boundaries at the brain edges. Supra-threshold TCM voxels will generally appear in clusters of a couple of voxels or more. The preponderance of the TCM voxels will appear in the inferior, frontal and temporal/lateral portions of the acquisition volume (outside the brain). Lateral sagittal slices typically exhibit a lot more TCM artifacts than medial sagittal slices.

3. Switch the underlay file to the MR angiogram and back to ensure that the TCM artifact is not corrupted by draining veins at the edges of the brain or on the skull.

4. Switch the underlay file to the 16-image deconvolution analysis' 3D + time impulse response dataset. Examine the temporal evolution of the first 5 points of the response, for sharp fluctuations in signal (10% changes or more over 2–3 s). Often large changes in signal will be followed by slow return to baseline as shown in Figure 2.

5. The biggest change in the signal will occur in the 2–3 images (time-points) of the TCM IRF, but sometimes onset of large signal changes is delayed by one or two images (time-points).

6. The shapes of TCM IRFs will be similar within a cluster but may vary across different clusters.

7. TCM IRFs can look similar in different slices. Visually examine the subsequent selected TCM IRFs to guard against duplication of representative TCM IRFs, which may increase computational load. However, every effort should be made to pick all distinctly shaped TCM IRFs. Generally, 10–15 distinct representative TCM IRFs are sufficient to classify all TCM corrupted voxels in the dataset.

8. For choosing representative BOLD HDRs, examine supra-threshold voxels in regions known to activate in the language FMRI paradigm. Regions expected to exhibit neural activation under the semantic overt word generation with auditory cues include, primary and secondary auditory cortex, superior temporal gyrus, Broca's area and lateral frontal cortex, inferior frontal gyrus motor cortex, premotor cortex, supplementary motor area and medial frontal cortex, thalamus, basal ganglia, etc.

9. The chosen BOLD HDRs should be slowly varying (i.e., should not exhibit large variations in signal from one time-point to the next).

10. As the early phases of the BOLD HDRs are most likely to overlap with TCM signal changes, care must be taken to sample representative BOLD HDRs that span all observed delays to onset and times to peak of putative BOLD voxels in the dataset.

11. BOLD HDR shapes typically do not vary much across regions and a few (3–5) representatives are probably sufficient to enable proper classification of all BOLD voxels in the dataset, i.e. the quantity CCB (obtained from these representatives) remains high for all putative BOLD voxels in the dataset.

APPENDIX B

List of AFNI Commands Used in Selective Detrending

The following is a list of AFNI commands used in the data analysis for selective detrending

1. 3dvolreg: for intensity-based 3D volume registration of spiral FMRI datasets; e.g., 3dvolreg −base overt5.unreg+orig[160]' −1Dfile overt1_mot.1D −prefix overt1.reg overt1.unreg+orig.

2. 3dTcat: for removing linear drifts followed by concatenation of the 5 FMRI runs. e.g., 3dTcat −rlt++ −prefix overt overt*BRIK

3. 3dDeconvolve: 1st deconvolution analysis, to estimate voxelwise IRFs from the input semantic word generation OAV and TCM-corrupted 3D+805 FMRI time-series dataset; e.g., 3dDeconvolve −input overt+orig −polort 0 −num_stimts 1 −stim_file 1 wordgen_OAV.1D −stim_minlag 1 0 −stim_maxlag 1 15 −iresp 1 overtML0.resp −fout −rout −nocout overtML0.buck.

4. Choosing representative TCM IRFs and BOLD HDRs (Appendix A).

5. Custom-built selective detrending Matlab program *seldet.m*: outputs a TCM artifact reduced FMRI time-series 'overt.seldet+orig'.

6. 3dDeconvolve: 2nd deconvolution analysis, to generate artifact reduced impulse response files and R^2 and F-statistic maps. e.g., 3dDeconvolve −input overt.seldet+orig −polort 0 −num_stimts 1 −stim_file 1 wordgen_OAV.1D -stim_minlag 1 0 −stim_maxlag 1 15 −iresp 1 overt.seldet.resp −fout −rout −nocout overt.seldet.buck.

7. 3dAutomask & 3dcalc: programs used to create extra- and intra-cerbral brain mask e.g. 3dAutomask −nograd −dilate 1 −prefix brain_mask overt.seldet+ orig 3dcalc

−prefix extra_cereb_mask −a brain_ mask+orig −expr '1-step(a)' e.g. 3dAutomask −nograd −erode 1 −prefix intra_cereb_mask overt. seldet+orig

APPENDIX C

Glossary of Terms

TCM: Task-correlated motion

HDR: Hemodynamic response

IRF: Impulse response function

ISI: Inter-stimulus interval

OAV: Overt-response locked analysis vector, Created by the temporal sequence of spoken words generated by the subject.

CCT: Absolute maximum of the cross-correlation coefficients between a given voxel's IRF and each of the 10–15 representative TCM IRFs

CCB: Maximum of the cross-correlation coefficients between a given voxel's IRF and each of the 3–5 representative BOLD HDRs.

MPR: Motion parameter regression

'dtsel': Selective detrending method

'dtr': Non-selective detrending method

'ML2': method of ignoring two images after speech. Obtained by setting the stim_minlag parameter in AFNI's 3dDeconvolve command to 2.

'ML0': uncorrected TCM artifact-ridden dataset, obtained by setting the stim_minlag parameter in AFNI's 3dDeconvolve command to 0.

α: Least squares linear regression estimate of the proportionality constant in the model: voxel IRF = $\alpha \times$ maximally correlated TCM IRF.

GLM: General linear model

TR/TE/FA: Repetition time/Echo Time/Flip angle

FOV: Field of View

MRA: Magnetic resonance angiogram

SPGR: Spoiled gradient recalled

3D: 3-dimensional

BOLD: Blood oxygen level dependent

FMRI: Functional magnetic resonance imaging

AFNI: Analysis of Functional Neuroimages