EVALUATION OF MULTI-ECHO ICA DENOISING FOR TASK BASED FMRI STUDIES: BLOCK DESIGNS, RAPID EVENT-RELATED DESIGNS, AND CARDIAC-GATED FMRI

[1]Javier Gonzalez-Castillo, [1]Puja Panwar, [1]Laura C. Buchanan, [2]Cesar Caballero-Gaudes, [1]Daniel A. Handwerker, [1]David C. Jangraw, [3]Valentinos Zachariou, [4]Souheil Inati, [4]Vinai Roopchansingh, [4]John A. Derbyshire, [1,4]Peter A. Bandettini.

[1]Section on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD.
[2]Basque Center on Cognition, Brain and Language, San Sebastian, Spain.
[3]Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD.
[4]Functional MRI Core, National Institute of Mental Health, National Institutes of Health, Bethesda, MD.

ABSTRACT

Multi-echo fMRI, particularly the multi-echo independent component analysis (ME-ICA) algorithm, has previously proven useful for increasing the sensitivity and reducing false positives for functional MRI (fMRI) based resting state connectivity studies. Less is known about its efficacy for task-based fMRI, especially at the single subject level. This work, which focuses exclusively on individual subject results, compares ME-ICA to single-echo fMRI and a voxel-wise T_2^* weighted combination of multi-echo data for task-based fMRI under the following scenarios: cardiac-gated block designs, constant repetition time (TR) block designs, and constant TR rapid event-related designs. Performance is evaluated primarily in terms of sensitivity (i.e., activation extent, activation magnitude, percent detected trials and effect size estimates) using five different tasks expected to evoke neuronal activity in a distributed set of regions. The ME-ICA algorithm significantly outperformed all other evaluated processing alternatives in all scenarios. Largest improvements were observed for the cardiac-gated dataset, where ME-ICA was able to reliably detect and remove non-neural T_1 signal fluctuations caused by non-constant repetition times. Although ME-ICA also outperformed the other options in terms of percent detection of individual trials for rapid event-related experiments, only 46% of all events were detected after ME-ICA; suggesting additional improvements in sensitivity are required to reliably detect individual short event occurrences. We conclude the manuscript with a detailed evaluation of ME-ICA outcomes and a discussion of how the ME-ICA algorithm could be further improved. Overall, our results suggest that ME-ICA constitutes a versatile, powerful approach for advanced denoising of task-based fMRI, not just resting-state data.

KEYWORDS: multi-echo fMRI, ME-ICA, rapid event related, block design, sensitivity

INTRODUCTION

Even with advances in hardware (e.g., higher field systems) and acquisition technology (e.g., multichannel receiver, surface coils, etc.), in functional magnetic resonance imaging (fMRI) there exists a continual need for a greater signal-to-noise ratio, especially at the single subject level. In addition to the signal of interest—T_2* fluctuations of a neuronal origin—fMRI time series contain fluctuations due to thermal noise, hardware instabilities, subject head motion, cardiac function and respiration (see (Greve et al., 2013) for a detailed review). In some instances, these nuisance sources can account for up to 82% of the variance at the voxel level (Bianciardi et al., 2009). Insufficiently accounting for these undesired sources of fluctuation during analyses translates into reduced sensitivity to true neuronal responses (Gonzalez-Castillo et al., 2012b), lower test-retest reproducibility, biased results across populations, and ultimately obstructs the interpretability of the results and diminishes their potential scientific and clinical value.

To reduce noise in fMRI data, complex pre-processing pipelines precede activation and connectivity analysis. For example, slow signal drifts are often modeled via Legendre polynomials or sinusoidal basis sets (see (Tanabe et al., 2002) for details). All fMRI analysis packages include tools to estimate and minimize artifacts from head motion (e.g., *3dvolreg* in AFNI, *mcflirt* in FSL). Several algorithms have been proposed to also moderate signal variance due to cardiac and respiratory function (e.g., RETROICOR (Glover et al., 2000), RVT (Birn et al., 2008), HR variation (Chang et al., 2009)). Spatially uncorrelated noise is often lowered by means of spatial smoothing. Finally, gains in signal-to-noise ratio can be obtained by averaging across runs and subjects. More advanced denoising methods include the use of multivariate decomposition approaches, such as principal component analysis (PCA) or independent component analysis (ICA), to identify and subsequently remove artifactual (i.e., non-neuronal) signals specific to each dataset. In the past, these procedures have relied on the expertise of well-trained fMRI specialists to manually identify noise components. For ICA, automatic classification of nuisance ICA components based on different combinations of spatial, temporal and spectral characteristics of the components have been recently proposed—namely FIX (Salimi-Khorshidi et al., 2014) and AROMA (Pruim et al., 2015))—yet, they either require a study-specific data-intensive training phase (i.e., FIX) or focus solely on a subset of noise sources (i.e., AROMA deals primarily with motion-related artifacts).

An alternative way to improve the sensitivity to the BOLD response, and in turn improve the contrast-to-noise of fMRI experiments, is to acquire the data differently, by using multi-echo acquisition schemes(Gowland and Bowtell, 2007; Poser et al., 2006; Posse, 2012; Posse et al., 1999; Speck and Hennig, 1998). In single-echo fMRI, data is acquired at a unique echo time (TE) close to the average grey matter T_2^* inside regions targeted by the study. Conversely, in multi-echo (ME) fMRI, the scanner outputs N_e time series per voxel, each of them acquired at a different TE, and all of them following a single excitation pulse. Crucially, these N_e time series

differ from each other in terms of T_2^* weighting and thermal noise, but not in terms of T_1 weighting. These specific properties of ME-fMRI data can be exploited in several manners for denoising purposes. For example, different voxel-wise linear weighted combination schemes of ME time series have been demonstrated to improve sensitivity for task and resting experiments at 1.5T (Posse et al., 1999), 3T (Poser et al., 2006) and 7T (Poser and Norris, 2009); primarily by reducing thermal noise and susceptibility artifacts. In addition, other groups have proposed dual-echo approaches where signal fluctuations recorded at a short echo (assumed to have minimal T_2^* weighing) are regressed out from the time series acquired at a longer echo time optimized for BOLD weighting (Bright and Murphy, 2013; Buur et al., 2009). One such example is the work of Bright and Murphy (2013) who evaluated how regressing out data acquired at TE=3.3 ms (expected to have minimal T_2^* weighting) from that acquired at TE=35 ms (a common TE for experiments at 3T) could help reduce motion-related effects and physiological noise. This approach showed improvements in connectivity estimations, but resulted in reduced activation extent and magnitude; therefore demonstrating that activity and connectivity studies may benefit differently from similar ME-based denoising schemes.

Kundu et al. (2012) recently proposed a ME-based denoising technique named ME-ICA (Multi-Echo Independent Component Analysis); which takes advantage of the distinct TE-dependence profiles of BOLD-like (linear dependence with TE) and non-BOLD-like fluctuations (no dependence with TE) to automatically classify ICA components as signal (BOLD-like) or noise (non-BOLD-like). The ME-ICA algorithm proceeds as follows. First, voxel-wise estimates of T_2^* are obtained. These are subsequently used to linearly mix all echoes and create a new single, "Optimally Combined", time series per voxel (the OC time series) optimized for functional contrast (Poser et al., 2006; Posse et al., 1999). This OC time series constitutes the input to a subsequent ICA that extracts spatially independent signals in the data. The overall TE-dependence profile of each ICA component is then characterized using two summary metrics: kappa (κ) and rho (ρ), respectively representing the BOLD signal and the spin-density or inflow signal. A combination of low kappa and high rho indicates the component has a low dependence on TE and a high likelihood of being noise (i.e., non-BOLD). In its last step, ME-ICA uses kappa, rho, explained variance and additional metrics that further characterize the TE-dependence profile of each component to automatically identify and regress out from the data those ICA components that constitute noise. (Kundu et al., 2012) and Appendix A in Olafsson et al. (2015) provide a detailed description of the ME-ICA algorithm.

Several studies have already established experimentally how ME-ICA can help improve the quality of fMRI results, but often focused on connectivity analyses. For example, ME-ICA has been shown to improve network detection over conventional single-echo fMRI both in humans at 3T (Kundu et al., 2012) and rats at 11.7T (Kundu et al., 2014). Olafsson and colleagues (2015) have also shown how ME-ICA can reliably identify and remove artifacts unique to novel simultaneous multi-slice acquisition techniques (Feinberg et al., 2010) during rest scans. In addition, ME-

ICA's ability to differentiate artifactual slow signal drifts from those of a BOLD origin has also been demonstrated with long (> 1 min) blocks of visual stimulation (Evans et al., 2015). However, little is known about the performance of ME-ICA for conventional task experimental designs. To the best of our knowledge, only one non peer-reviewed study to date has evaluated ME-ICA for regular task-based fMRI (Lombardo et al., 2015). That study focused solely on sensitivity for block designs at the group level, concluding that ME-ICA outperforms single-echo fMRI and optimal combination of ME datasets in this particular scenario. Here, we attempt to address this gap by evaluating the performance of ME-ICA for task-based fMRI at the single-subject level. We decided to focus on single-subject results, as this is where fMRI holds potential clinical value, and where fMRI strives the most for improvements in signal-to-noise.

Here we evaluate ME-ICA with respect to single-echo fMRI and OC using both block and rapid event-related designs. For the block design evaluation we used an auditory task and two different acquisition strategies: constant-TR, and cardiac-gated (i.e., non-constant TR) acquisitions. Cardiac-gated fMRI constitutes an interesting test case for ME-ICA, as it produces data strongly contaminated by baseline signal fluctuations of T_1 origin (due to the non-constant TR) that ME-ICA should be able to reliably correct. For this particular scenario, we also evaluate ME-ICA against two additional processing strategies (Beissner et al., 2010; Guimares et al., 1998) previously shown to benefit in the analysis cardiac-gated datasets.

For the rapid event-related evaluation we acquired data for five different tasks (e.g., motor, auditory, reading, and two visual identification tasks), so that we could evaluate, to a given extent, the generality of ME-ICA performance across tasks. In all instances, we evaluate performance in terms of activation extent, activation magnitude, and effect size. We also investigate how ME-ICA can help detect activity for individual instances of tasks (i.e., individual trials). Our results show how ME-ICA significantly outperforms the single-echo fMRI and OC pipelines in all scenarios, suggesting that task-based fMRI can benefit from the ME-ICA approach just as resting-state connectivity analyses have done in the past.

METHODS

We acquired two different datasets for this study. First, we collected fMRI data from five individuals using an auditory block-design paradigm. For this first dataset, we acquired data in two ways: constant repetition time (non-gated) and cardiac-gated. Second, we collected fMRI data from ten additional subjects using a multi-task rapid event-related paradigm. This second dataset was acquired using a constant repetition time only. Acquisition details, analytical procedures and experimental goals for both datasets are described below.

Block Design Experiments

Subjects. Five subjects (2 males, 3 females, mean ± SD age = 25 ± 2 y.o.) participated in these experiments after giving informed consent in compliance with the NIH Combined Neuroscience Institutional Review Board-approved protocol 93-M-0170 in Bethesda, MD.

Experimental Paradigm. All functional scans (cardiac-gated and non-gated) were acquired using the same block design paradigm. An initial 20 seconds period was followed by 5 repetitions of the following sequence of blocks: listen block (20 s); and rest block (40 s). An additional 20 seconds of rest were added at the end of each functional run. This resulted in 340-second runs. During the rest periods, subjects were instructed to remain still and focus their attention on a crosshair at the center of the screen. During the listen blocks subjects were presented with 20 s of instrumental music via MRI compatible headphones. Subjects had been previously instructed to attentively listen to the music when present. The *PsychoPy* software (Peirce, 2008) was used for stimuli delivery.

Data Acquisition. Imaging was performed on a General Electric (GE) 3 Tesla 750 MRI scanner (Waukesha, WI). The scanner's body coil was used for RF transmission, and a 32-channel receive-only head coil (GE, Waukesha, WI) was used for signal reception. Functional scans were acquired with a multi-echo EPI sequence (flip angle=60°, TEs=13.9/31.7/49.5 ms, 33 oblique slices, slice thickness=3 mm, in-plane resolution=3x3 mm^2, FOV 216 mm, acceleration factor 2, number of acquisitions = 136, bottom/up sequential acquisitions). For each participant we acquired two non-gated functional scans and two cardiac-gated functional scans. The order of the scans was randomized across subjects. Non-gated scans were acquired using a constant repetition time (TR) of 2.5 seconds. Cardiac-gated acquisitions were time-locked to the first peak of the cardiac cycle, recorded on a GE optical pulse oximeter attached to one of the subject's fingers, following a nominal TR of 2.5s. This resulted in a non-constant TR of mean ± SD = 3.12 ± 0.15 seconds across the whole dataset. At the end of each cardiac-gated scan, the system saved a text file with information about the actual repetition time between successive acquisitions. Evaluation of cardiac traces and triggering files confirmed reliable detection of cardiac cycle events, and correct synchronization of fMRI triggering events with the peak of the cardiac cycle.

In addition, T_1-weighted Magnetization-Prepared Rapid Gradient-Echo (MPRAGE) and Proton Density (PD) sequences were acquired for presentation and alignment purposes (axial prescription, number of slices per slab, 176; slice thickness, 1 mm; square FOV, 256 mm; image matrix, 256 × 256).

Data Pre-Processing. Data were pre-processed with the AFNI software (Cox, 1996). Three different pre-processing pipelines were used in these experiments, namely: single-echo (1E) pipeline, OC pipeline, and ME-ICA pipeline. In addition, for the cardiac-gated dataset, data were pre-processed with an additional single-echo pipeline that included a model-based correction of T_1 baseline signal fluctuations associated with non-constant TRs following the procedures previously described by

Guimares et al. (1998). This additional cardiac-gated only pipeline is referred to as the 1E-T_1C (after T_1 Correction) pipeline throughout the manuscript.

1E Pipeline. Only one time series, that for TE=31.7 ms, enters this pipeline, in a manner similar to how single-echo fMRI is commonly analyzed in conventional fMRI studies. The pre-processing steps in the 1E pipeline are: (1) discard initial 10 seconds of data to achieve steady-state, (2) time-shift correction, (3) estimation of head motion (AFNI program *3dvolreg*) and transformations to MNI space using the MPRAGE and PD scans following procedures previously described in (Gonzalez-Castillo et al., 2012a), (4) spatial smoothing (FWHM = 6 mm; AFNI program *3dBlurInMask*), and (5) voxel-wise intensity normalization to percent signal change units.

For non-gated data, the time-shift correction step was performed with AFNI program *3dTshift*, which interpolates data in time so that all slices have the same temporal reference. To accomplish this, *3dTshift* takes as input information about slice acquisition order. For cardiac-gated data, time-shift correction was performed with AFNI program *3dTRfix*, which not only corrects inter-slice timing, but also brings non-constant TR datasets into a regular temporal grid; also via linear interpolation in time. *3dTRfix* takes as inputs not only the slice acquisition order, but also the onsets of each volume acquisition. These two ways to time-shift correct the data are common across all other pre-processing pipelines described below.

1E-T_1C Pipeline. This pipeline also takes only the second echo (TE=31.7 ms) as input. The only difference from the 1E pipeline, is that a model-based T_1 baseline shift correction following procedures previously described by Guimares et al. (1998) was performed after discarding the initial 10 seconds of data, and prior to all other pre-processing steps. No other differences exist between the 1E and 1E-T_1C pipelines.

T_2^* Estimation Pipeline. This pipeline uses the last two echoes (31.7ms and 49.5ms) and follows methods previously proposed by Beissner et al. (2011, 2010) for the analysis of non-constant TR fMRI datasets. Pre-processing steps include: (1) discard initial 10 seconds of data, (2) time-shift correction (*3dTshift* or *3dTRfix*), (3) estimation of head motion and transformation to MNI space using the MPRAGE and PD scans following procedures previously described in (Gonzalez-Castillo et al., 2012a), (4) spatial smoothing (FWHM = 6mm), (5) computation of T_2^* time series using Eq. 1, and (6) voxel-wise intensity normalization to signal percent change units.

$$T_2^*(i,n) = \frac{TE_3 - TE_2}{ln(S_2(i,n)/S_3(i,n))} \quad \text{(Eq. 1)}$$

In Eq. 1 (originally derived in (Beissner et al., 2010)), index *i* represents voxel, index *n* represents time, TE_2 and TE_3 refer to the two experimental echo times, S_2 refers to the time series for TE_2=31.7ms, and S_3 refers to the time series for TE_3=49.5ms.

OC Pipeline. All three echoes are used in this pipeline. Pre-processing steps include: (1) discard initial 10 seconds of data, (2) time-shift correction, (3) estimation of head motion and transformations to MNI space using the MPRAGE and PD scans following procedures previously described in (Gonzalez-Castillo et al., 2012a), (4) voxel-wise linear weighted combination of echoes optimized for T_2^* (Poser et al., 2006; Posse et al., 1999) with AFNI program *tedana.py* (a component of the ME-ICA software; version 2.5-beta11; https://bitbucket.org/prantikk/me-ica#f5d52a6), (5) spatial smoothing (FWHM = 6 mm), and (6) voxel-wise intensity normalization to signal percent change units.

The voxel-wise weights for the OC time series are given by

$$w_{i,v} = TE_i \cdot e^{-TE_i / \widehat{T_2^*}} \quad \text{(Eq. 2)}$$

where i=1..3 refers to echo, v refers to voxel, and $\widehat{T_2^*}$ corresponds to voxel-wise estimates of T_2^* obtained via a log-linear firs to the multi-echo dataset.

ME-ICA Pipeline. All three echoes are used in this pipeline. This pipeline builds on top of the OC pipeline. Following step (4) of the OC pipeline, we use the AFNI program *tedana.py* (*version 2.5, beta 11*) to perform ME-ICA denoising (Kundu et al., 2012). The denoised time series output by ME-ICA are then submitted to (5) spatial smoothing (FWHM = 6 mm), and (6) voxel-wise intensity normalization steps, as in the 1E and OC pipelines.

Activation Analyses. Following each pre-processing pipeline, activation maps for the contrast task vs. rest were obtained separately for each subject using both functional runs as input to AFNI program *3dREMLFit.* Motion parameters, their first derivatives and Legendre polynomials up to 3rd order were used as covariates of no interest. All activation maps were thresholded at $p_{FDR}<0.05$.

For the ME-ICA pipeline, degrees of freedom were adjusted to account for the number of removed components when computing statistical significance for this particular pipeline. One degree of freedom is subtracted for each component being removed (Kundu et al., 2012).

Target Regions of Interest (ROI). Performance metrics for this first set of experiments were computed using two different sets of ROIs: (1) ROIs derived from the reverse inference map generated by the *Neurosynth* tool (Yarkoni et al., 2011) for the concept "music"; and (2) bilateral inferior colliculus (IC) ROIs based on previously published coordinates (described below).

[INSERT FIGURE 1 APROX. HERE]

The *Neurosynth*-derived ROIs (Fig. 1.A) include 4 ROIs covering primarily bilateral superior temporal cortex and bilateral pre-central gyrus. The IC ROIs (Fig. 1.B) are 5

mm radius spherical ROIs centered at reference MNI coordinates for this particular anatomical structure (*[x,y,z]=[±6,-33,-9]*) taken from (Parsons et al., 2014). We decided to evaluate the IC region, as it is a well established processing node of the ascending auditory pathway in which detection of activity is difficult due to large amounts of pulsatile noise (Guimares et al., 1998).

Performance Metrics. The different pre-processing pipelines are evaluated in terms of activation extent, T-statistic magnitude, and estimated effect size.

Activation extent was measured as the number of significantly active voxels inside each set of target ROIs (e.g., *Neurosynth* and IC ROIs). Activation extent measures for each pipeline were included in a *2-way mixed-effects ANOVA [A=Subject|Random; B=Pipeline|Fixed]* in MATLAB to elucidate if pre-processing pipelines had a significant effect on activation extent. This omnibus test was followed by post-hoc paired *T-tests* to discover significant pair-wise differences across pipelines. For non-gated data we only compared three pipelines—namely 1E, OC and ME-ICA—while all 4 pipelines were compared using cardiac-gated data.

T-statistic magnitudes were used as a proxy for contrast-to-noise for the condition of interest. Average *T-statistic* values were computed within significantly active voxels inside target ROIs. For all pipelines, we computed averages using only the voxels that were significant during the *1E* pipeline to ensure consistency of voxels contributing to the averages. Similar to activation extent, average *T-statistics* were submitted to an *ANOVA*. When a main effect was found, post-hoc *paired T-tests* across pipelines were conducted.

Finally, effect size was evaluated in terms of average coefficient fits (i.e., beta weights) within significantly active voxels inside the target ROIs. The same statistical tests used for the other two metrics were also used for effect size measures.

Estimates of T_1 baseline fluctuations in cardiac-gated data. To evaluate how well ME-ICA identifies this artifact specific to the cardiac-gated dataset, we computed the temporal Pearson correlation between time series of components marked as noise by the ME-ICA algorithm and estimates of T_1-related baseline shift estimated using the following equation:

$$T_1 Shift(t) = 1 - e^{-TR(t)/T_1} \quad \text{(Eq. 3)}$$

where *TR(t)* refers to the time between onsets of consecutive volume acquisitions; and T_1 was set to 1331 ms, according to previous estimates of T_1 for grey matter at 3T published by Wansapura et al. (1999). This equation is equivalent to equation 1 in Guimares et al. (1998).

BOLD Contrast Simulation. To understand how OC affects effect size estimates (i.e., BOLD contrast estimates), we generated theoretical BOLD contrast and OC weight curves for TEs ranging from 0ms to 200ms (Fig. 2.A). OC weight curves were generated using Eq. 2 above. BOLD contrast curves were generated using equation 3 from Posse et al. (1999), reproduced here:

$$\Delta S(TE_i) = S_o \frac{TE_i}{T_2^*} e^{(-TE_i/T_2^*)} \frac{\Delta T_2^*}{T_2^*} \text{ (Eq. 4)}$$

where i indexes echo time, So is the average initial signal amplitude, T_2^* is the average transverse relaxation time due to spin-spin interactions and static field inhomogeneities, and ΔT_2^* is the stimulus dependent change in T_2^* relaxation time. So and T_2^* were estimated directly from the data using AFNI program *tedana.py* (version 2.5, beta 11), which generates voxel-wise estimates of So and T_2^*. For the purpose of these simulations, we computed a single So and T_2^* value as the average of the corresponding voxel-wise maps within significantly active voxels for the non-gated block-design experiments across all subjects (T2*= 44.3 ± 3.6; So= 1978 ± 109). BOLD contrast curves were generated for three different ΔT_2^* scenarios: 1, 3 and 7 percent changes in T_2^*.

Once these curves were available, we extracted estimates of BOLD contrast (ΔS) and OC weights (w_i) for our three experimental TEs. These were subsequently used to compute theoretical estimates of BOLD contrast for the OC approach. Figure 2.B shows the estimated BOLD contrast for the middle echo (red bars) and OC (green bars) for the different ΔT_2^* scenarios. In all instances, the OC BOLD contrast is lower than the middle echo BOLD contrast; which predicts a decrease in estimates of effect size for the OC pipeline relative to the 1E pipeline should be observed in the experimental results.

[INSERT FIGURE 2 APROX. HERE]

Rapid Event-related Experiments

Subjects. Ten subjects (5 males, 5 females, mean ± SD age = 25 ± 3 y.o.) participated in these experiments after giving informed consent in compliance with the NIH Combined Neuroscience Institutional Review Board-approved protocol 93-M-0170 in Bethesda, MD.

Experimental Paradigm. The *PsychoPy* software (Peirce, 2008) was used for stimulus delivery. Eye tracking data were collected to check subject's performance (see below for further description). Subjects were instructed on five different tasks prior to entering the scanner room. The purpose of using five distinct tasks was to engage multiple cognitive systems in a single event-related study to make sure the effects of pre-processing choices were not restricted to specific brain regions. The tasks used in these experiments are:

401
402 [INSERT FIGURE 3 APROX. HERE]
403
404 • *Motor (MOTOR).* Subjects were instructed to intermittently press one button of a
405 response box with a single finger at a fixed rate of approximately 0.5Hz. By using
406 the response box we were able to objectively evaluate subject compliance for
407 this task. Motor task trials always lasted 4 seconds. During these trials, three
408 items were presented on a screen (Figure 3.A): a central crosshair to aid with
409 fixation, a left pointing arrow that didn't relate to task instructions, and an
410 integer counter to help subjects press the button at a constant rate. All subjects
411 performed this task with the left hand except two, who were inadvertently
412 provided with the response box on their right hand. The hand used during the
413 task was taken into account during the analyses.
414
415 • *Biological Motion Observation (BMOT).* Subjects were instructed to observe short
416 4-second publicly available videos of dot patterns resembling biological motion
417 such as walking, jumping, dancing, drinking and climbing steps. During
418 biological motion task trials, the crosshair disappeared from the center of the
419 screen and the corresponding video appeared on one of the two visual hemi-
420 fields (right or left; Figure 3.B). The position of the videos was randomized
421 across trials to aid with eye tracking data analysis.
422
423 • *Passive Viewing of Houses (HOUSES).* Subjects were instructed to attentively look
424 at a succession of pictures of houses that appeared in the center of the screen
425 (Figure 3.C). Each house task trial lasted 4 seconds, during which subjects were
426 presented with six different houses. Each house appears for approximately 170
427 milliseconds with a gap of approximately 500 milliseconds in between totaling
428 to the 4 seconds per trial.
429
430 • *Listening to Music (MUSIC).* Subjects were instructed to attentively listen to 4-
431 second recordings of music clips played by a single instrument—namely violin,
432 piano or drums—and to direct their gaze to the picture representing the
433 instrument being played as soon as they had identified it. During each music
434 trial, in addition to the auditory stimuli, subjects were presented with three
435 pictures (one per instrument) located in a triangular arrangement on the screen
436 (Figure 3.D).
437
438 • *Sentence Reading (READ)*. Subjects were instructed to covertly read sentences
439 presented on the screen one word at a time (Figure 3.E). For each trial, words
440 were presented in one of the two hemifields (right or left) to aid with analysis of
441 eye tracking data. All words of a trial appeared on the same hemifield. Each word
442 was presented for approximately 250 milliseconds, with gaps of approximately
443 100 milliseconds in between. Sentences ranged in length between 10 and 11
444 words, so each trial lasted either 3400 or 3750 milliseconds.
445

All functional runs lasted 440 seconds and contained a total of 30 task trials, six per task type. Onset times for trials were obtained with *Freesurfer* program *optseq2* (*https://surfer.nmr.mgh.harvard.edu/optseq/*), which is designed to optimize timing of events for event-related experiments. Three different schedules (onset times) were randomly used in these experiments. For all three schedules the minimum inter-stimulus interval (ISI) was 10 s. Mean and standard deviation ISIs for the three different schedules were: 13 ± 24, 13 ± 18 and 13 ± 15 seconds.

MRI Data Acquisition. Imaging was performed on a General Electric (GE) 3 Tesla 750 MRI scanner (Waukesha, WI). The scanner's body coil was used for RF transmission, and a 32-channel receive-only head coil (GE, Waukesha, WI) was used for signal reception. Functional scans were acquired with a multi-echo EPI sequence (flip angle=70° for 9 subjects, flip angle=60° for 1 subject, TEs=16.3/32.2/48.1 ms, TR=2 s, 30 axial slices, slice thickness=4 mm, in-plane resolution=3x3 mm^2, FOV 192 mm, acceleration factor 2, number of acquisitions = 220, bottom/up sequential acquisitions). For one subject, acquisitions were interleaved, instead of sequential due to an operator error. Two functional runs were acquired in six subjects, and only one in the remaining four due to scanning time constraints.

In addition, MPRAGE and PD sequences were acquired for presentation and anatomical alignment purposes (axial prescription; number of slices per slab, 176; slice thickness, 1 mm; square FOV, 256 mm; image matrix, 256 × 256).

MRI Data Pre-processing. Data were also pre-processed with the AFNI software (Cox, 1996) using the three main pipelines described above: 1E pipeline, OC pipeline and ME-ICA pipeline. All runs in this second experiment were acquired using a constant TR, therefore the slice time correction step is always performed with AFNI program *3dTshift*.

Activation Analysis. For this second experiment, each functional run was analyzed separately. Two different statistical analyses were conducted to generate activation maps. First, we computed activation maps per task-type, taking into account all 30 trials in a run. Second, we attempted detection of individual trials, generating activation maps for each trial independently.

For the ME-ICA pipeline, degrees of freedom were adjusted to account for the number of removed components when computing statistical significance for this particular pipeline.

Per-Task Activation Maps. Following each pre-processing pipeline, activation maps for the five contrasts of interest (e.g., music vs. rest, read vs. rest, etc.) were obtained separately for each run and subject using AFNI program *3dREMLFit.* Motion parameters, their first derivatives and Legendre polynomials up to 3rd order were used as covariates of no interest. All activation maps were thresholded at $p_{FDR}<0.05$.

Per-Trial Activation Maps. Following each pre-processing pipeline, activation maps for each individual trial of each task were also computed with AFNI program *3dREMLFit*, using the individual modulation option (-stim_times_IM) that instructs the program to generate a separate regressor per individual trial so that it can compute statistics (i.e., effect size and T-stat) for each individual task event. Motion parameters, their first derivatives and Legendre polynomials up to 3rd order were used as covariates of no interest. All activation maps were thresholded at $p_{FDR}<0.05$.

Target Regions of Interest. Five different target ROI sets, one per task type, were used in this second experiment. Three ROI sets (i.e., HOUSES, MUSIC and READ) were obtained using the *Neurosynth* tool (Yarkoni et al., 2011) and two (i.e., MOTOR and BMOT) using cytoarchitectural maximum probability maps distributed with AFNI.

- *MOTOR ROIs.* Voxels with a probability of being in task-contralateral anterior and posterior Brodmann Area 4 greater than 70% according to cytoarchitectural maximum probability maps for these regions (Geyer et al., 1996) distributed with AFNI (Figure 4.A).

- *BMOT ROIs.* Voxels inside bilateral human occipital visual area 5 (hOC5) according to cytoarchitectural maximum probability maps for this region (Malikovic et al., 2006) distributed with AFNI (Figure 4.B).

- *HOUSES ROIs.* Voxels inside reverse inference maps generated by the *Neurosynth* tool for the concept "*place*" covering bilateral place parahippocampal region and bilateral posterior cingulate cortex (Figure 4.C). No map for concept "house" is currently available at *Neurosynth*.

- *MUSIC ROIs.* These are the same ROIs derived from *Neurosynth* used in experiment one.

- *READ ROIs.* Voxels inside reverse inference maps generated by the *Neurosynth* tool for the concept "*reading*" covering primarily large portions of left inferior and middle frontal gyrus, the left posterior superior temporal gyrus, the left fusiform gyrus (e.g., visual word form area) and bilateral higher visual regions (Figure 4.D).

[INSERT FIGURE 4 APROX. HERE]

Performance Metrics. Similarly to the block-design experiments, we evaluated the different pre-processing pipelines in terms of activation extent, *T-statistic* magnitude and effect size using the per-task activation maps and task-specific target ROIs. We computed the three metrics for each task, and then evaluated if there was any significant difference across pre-processing pipelines using a *3-way mixed-effects ANOVA [A=Subject|Random; B=Pipeline|Fixed; C=Task|Fixed]* in MATLAB.

Post-hoc *paired T-tests* between pre-processing pipelines were also conducted to detect significant differences between pairs of pipelines.

For the per-trial analysis, we evaluated the performance of the different pre-processing pipelines in terms of the percentage of detected trials. Each run contains 30 trials (6 per task) leading to a total of 480 trials (*6 subjects* x *2 runs* x *5 tasks* x *6 events* + *4 subjects* x *1 run* x *5 tasks* x *6 events*) in these experiments. Subjects were actively engaged in 473 of them according to behavioral and eye tracking data (see details below). For each attended trial, we evaluated its fMRI activation map, and marked the trial as "detected" if its associated activation map had at least 10 significantly active voxels ($p_{FDR} < 0.05$) inside the appropriate target ROI set. Finally, percent of detected trials per task/per subject were computed and input to a *3-way mixed effect ANOVA* and subsequent post-hoc *paired T-tests* in a manner similar to all other performance metrics.

Temporal Signal-To-Noise Ratio (TSNR). To evaluate how much the percentage of detected events correlates with initial data quality, we computed the TSNR of all scans. TSNR is defined voxel-wise as the ratio of the temporal average of the signal for a given voxel divided by the temporal standard deviation of the signal in that same voxel (Parrish et al., 2000). Voxel-wise TSNR values are often averaged across voxels for reporting purposes.

Here we first computed voxel-wise TSNR maps for each echo separately right after discarding non-steady state volumes. A single TSNR value per echo *(e)* and functional run *(r)* was computed as the average of TSNR values across all intra-cranial voxels in the imaging field of view. Additionally an overall (across echoes) TSNR value per functional run *(r)* was computed using the following equation:

$$TSNR_r = \frac{\sum_{e=1}^{3} TE_e \times TSNR_{r,e}}{\sum_{e=1}^{3} TE_e} \quad \text{(Eq. 5)}$$

Eye Tracker Data Acquisition. An MRI compatible infrared eye tracking system, consisting of an infrared source and camera, mounted on top of the head coil was used to track the right eye's gaze position during functional scans (Avotec Real Eye Model RE-5701; Avotec Inc., Stuart, FL). The eye-gaze position was sampled at a rate of 60Hz using the SMI *iViewX* software. The *PsychoPy* software, in conjunction with home developed python add-ons for integration with *iViewX* (https://github.com/djangraw/PsychoPyParadigms), was used for running the 13-point eye tracker calibration protocol, as well as for synchronizing eye tracker recordings with stimulus presentation during functional scans.

Eye Tracking/Behavioral Data Analysis. The eye-tracking data were analyzed using the Open Gaze And Mouse Analyzer package (OGAMA; (Vosskühler et al., 2008)). Visual field areas of interest (AOI) specific to each task (Supplementary Figure 1) were first defined using OGAMA's AOI tool. We then used the *OGAMA Statistics Module* to compute fixation time and percent time inside target AOIs for

each trial. We declared events as *"valid"* (e.g., subjects attended to them) if the subject's total gaze duration inside the target AOI was equal or greater than 75% of the trial's duration. This cutoff at 75% was used to account for inherent jitter in the eye tracker and for blinking periods. For trials with large jitter, we used the OGAMA *Replay Module* to ensure the eye path showed a focus on the target AOI prior to assigning a definitive "valid" label.

For MOTOR trials, an event was declared "valid" only if in addition to fulfilling the above-mentioned eye-tracking criteria, a minimum of 4 button presses (half the number of expected presses) was logged by the response box during the trial. For four functional scans button box responses were not recorded due to an operator's error. Only the eye-tracking criterion was applied to these subjects.

For MUSIC trials, an event was declared "valid" only if in addition to fulfilling the above-mentioned eye-tracking criteria, subjects directed their gaze towards the instrument picture that corresponds to the actual instrument being played during that specific trial.

RESULTS

Block Design Experiments

[INSERT FIGURE 5 APROX. HERE]

Figure 5 shows probabilistic maps of activation (i.e., the color of a voxel indicates the percentage of subjects for which that voxel was marked as statistically significant at $p_{FDR}<0.05$) for the music vs. rest contrast for all pre-processing pipelines evaluated using the block-design dataset. Target ROIs are shown with a black contour. In all instances, strong activation was present in bilateral portions of the superior temporal gyrus, including, but not limited to, primary auditory cortex. Activity was present in additional regions only for a subset of the subjects. In all instances, ME-ICA produced the highest number of significantly active voxels (Figure 6.A). Regarding the IC, only the ME-ICA pipeline detected activity in this structure using cardiac-gated datasets (Table 1.A), and it did so for all 5 subjects. For non-gated data, although activity could be detected in some instances, ME-ICA had the best sensitivity (Table 1.B).

[INSERT TABLE 1 APROX. HERE]

Figure 6 shows quantitative results from the performance analyses conducted in the block-design datasets using the *Neurosynth*-derived target ROI set. A significant effect for the "pipeline" factor was found during the ANOVA analyses for activation extent ($F=82.15$; $p<0.05$), *T-statistic* magnitude ($F=104.75$; $p<0.05$) and effect size ($F=6.12$; $p<0.05$) for the non-gated data (top row in Figure 6). Subsequent paired T-tests between pairs of pipelines revealed that ME-ICA produced maps with

significantly more activation extent and higher *T-statistic* than all other pipelines (marked with an asterisk in Figure 6). For effect size, both OC and ME-ICA produced significantly smaller estimates than the 1E pipeline. In the case of cardiac-gated data (bottom row in Figure 6), we also included in the comparison the 1E-T1C pipeline (orange bars). For cardiac-gated data, we also found a significant main effect for pipeline in activation extent (F=89.6; p<0.05) and activation magnitude (F=74.2; p<0.05), but not in effect size (F=2.4; p=0.11). ME-ICA significantly outperforms all other pipelines in terms of activation extent and activation magnitude, according to post-hoc *paired T-tests* (asterisks in Figure 6). In addition, the 1E-T_1C pipeline resulted in significantly higher activation extent and T-stat than the 1E pipeline, but significantly less than ME-ICA.

Finally, both for gated and non-gated data, the T2* pipeline (yellow bars) resulted in significantly lower activation extent and magnitude than both the 1E and ME-ICA pipelines marked with asterisks in Figure 6.

ME-ICA reliably corrects T_1 baseline shifts in cardiac-gated datasets

Performance analyses for the cardiac-gated data suggest that ME-ICA was able to effectively identify and remove nuisance T_1-related baseline fluctuations associated with the non-constant TR. Figure 6 shows, for each subject, the one ME-ICA component with the highest rho (i.e., non-BOLD index). Such a component, which was always marked as noise by the algorithm, had in all cases an associated time series that correlated well (r = 0.84 ± 0.16) with estimates of T_1 baseline signal fluctuations obtained with Eq. 1 (red traces in the Figure 7). Moreover, the spatial maps of these components (right of Figure 7) resemble anatomical-like scans with good contrast between tissue compartments (e.g., grey matter, white matter and CSF). These are substantially different from the spatial maps of typical noise components associated with motion or scanner artifacts, as well as, from spatial maps typical of BOLD-like components (see Supplementary Figure 2 for representative components of each type).

Event Related Experiments

Figure 8 shows probabilistic maps of activation across runs for the per-task analyses in all five tasks for the three main pipelines under evaluation (1E, OC and MEICA). The color of a voxel in these maps indicates the percentage of runs for which that voxel was marked as significant (p_{FDR}<0.05). Per-task target ROI sets are shown as black contours. For all tasks, there is an increase in the extent of areas of high probability of activation (red arrows) going from left (1E pipeline) to right (ME-ICA pipeline). Such increases are not constrained solely to the target ROIs for each task, but also happen outside them (e.g., medial supplementary motor cortex for the motor task). In addition, figure 8 shows how there is an increase in the extent of significantly active regions for the multi-echo pipelines, especially for ME-ICA. This is particularly true for subcortical regions (black arrows).

Figure 9 shows the results for the different performance metrics for the per-task analyses. Similar to the block-design results, we observed an increase in activation extent and *T-statistic* magnitude for the ME-ICA pipeline with respect to OC and 1E pipelines. Conversely, the effect size estimate decreased for the OC and ME-ICA pipelines compared to the 1E pipeline. This is true for all five tasks. *ANOVA* revealed a main effect for pipeline in all three metrics: activation extent (F=90.3; p<0.05), activation magnitude (F=40.4; p<0.05), and effect size (F=86.9; p<0.05). Post-hoc *paired T-tests* revealed significant pair-wise differences between all pipelines for all tasks in terms of activation extent (marked with asterisks in Figure 9). For activation magnitude and effect size, a significant difference between ME-ICA and the 1E pipeline was detected in all instances. These results are consistent with those from the block design experiments and suggest that ME-ICA helps improve the sensitivity of rapid event-related experiments for a variety of tasks.

Finally, figure 10.A shows the percent of detected trials (p_{FDR} < 0.05) for all pipelines and tasks coming out of the per-trial analyses. The fraction of detected trials is below 50% for the MOTOR, BMOT and READING task in all instances. Only the HOUSES and MUSIC tasks reach levels above 50% for the ME-ICA pipeline. In fact, in all cases, ME-ICA produced on average a larger percent of detected events than the other two pipelines (except for the reading task, where the OC and ME-ICA pipeline have similar results). The *ANOVA* on this metric also showed a significant main effect for the factor "pipeline" (F=10.3; p<0.05), yet post-hoc *paired T-tests* only revealed a significant difference in percent of detected trials for the ME-ICA pipeline for the MOTOR, HOUSES and MUSIC tasks. When a less restrictive threshold (Figure 10.B, p_{Unc} < 0.001) is used, the number of detected events increases, yet the relationships between pipelines described above remain.

Detailed evaluation of individual subject results revealed substantial inter-subject differences in the percent of detected trials. While some subjects had relatively high detection percent for most tasks (e.g., above 50%), others had low detection percent (e.g., below 20%) for all tasks. Low detectability for individual events is expected given limitations in statistical power when attempting their detection on a trial-by-trial basis; yet to better understand such inter-subject differences, we computed TSNR for all available scans in the event-related dataset (Figure 11.A). We also computed the weighted average TSNR across all three echoes as a proxy for overall data quality of multi-echo scans (Figure 11.A, grey region). Figure 11B – D show scatter plots of percent-detected events versus TSNR for all three pipelines, excluding the scan marked as an outlier (Figure 11.A). In all instances, including ME-ICA, we observed a significant correlation between percent of detected trials and initial data quality (TSNR). This suggests that despite the ME-ICA denoising step there is still a strong dependence of the results on the original quality of the data.

Number of ME-ICA components.

Table 2 shows the average number of components found by the ME-ICA algorithm for all datasets, as well as the number of components being rejected (i.e., marked as noise by the algorithm).

DISCUSSION

Here we have evaluated the performance of ME-ICA with respect to an optimal linear weighted combination of multi-echo time series (Poser et al., 2006; Posse et al., 1999) and conventional analysis of single-echo fMRI data under different task-based experimental setups (i.e., block designs and rapid event-related designs) and acquisition strategies (i.e., constant TR and cardiac-gated). In all instances, ME-ICA outperformed the other pre-processing approaches in terms of activation extent, activation magnitude and ability to detect responses to individual trials. This, despite an average reduction in available degrees of freedom due to removal of noise ICA components (see Table 2). All together, these results suggest that ME-ICA can reliably help improve the sensitivity of task-based fMRI experiments at the single subject level. Thus, we extend prior work focused primarily on the evaluation of ME-ICA within the framework of resting-state and functional connectivity fMRI studies.

ME-ICA for cardiac-gated acquisitions

Cardiac-gated fMRI datasets constitute a particular case of fMRI data contaminated by strong T_1-related baseline signal fluctuations derived from non-constant acquisition times. As such, specific analytical techniques have been proposed for dealing with these particular datasets. Here, in addition to the 1E and OC pipelines, we also evaluated ME-ICA against two such specific methods: (a) a model-driven T_1 baseline signal correction method (1E-T_1C pipeline; (Guimares et al., 1998)); and (2) a dual-echo T_2^* estimate approach (T_2^* pipeline; (Beissner et al., 2010)). Our results show how ME-ICA outperformed all four pipelines in terms of activation extent and magnitude; suggesting ME-ICA may be the preferable approach for such datasets. Below we discuss the relative differences across all pipelines.

Only two pipelines, the 1E-T_1C and ME-ICA, significantly outperformed the 1E pipeline in cardiac-gated datasets. This suggests that these two pre-processing approaches were able to account, at least partially, for the above-mentioned T_1-related signal fluctuations. The same was not true for T_2^* and OC. For example, while the OC pipeline significantly outperformed the 1E pipeline for non-gated data, such was not the case for cardiac gating. In fact, for gated data, activation extent and magnitude significantly decreased for the OC pipeline relative to 1E (Figure 6). This is because the T_1 artifacts that affect gated acquisitions are equally present in all echoes, and a simple linear voxel-wise combination of the different echo time series does not eliminate them. Contrarily, it seems to enhance them relative to other signal components according to our results.

In the case of the T2* approach, this pipeline resulted in the lowest activation extent and magnitude for both gated and non-gated datasets. Although for the purpose of T_2^* estimation, data were spatially smoothed prior to computing T_2^* estimates (as previously suggested (Beissner et al., 2010)), numerical instabilities in the computation of the voxel-wise signal quotients (Eq. 1) might have resulted in T_2^* estimates with higher noise than original single echo time series; leading to the reported decrease in activation extent and magnitude. Similarly to here, a decrease in activation extent from single echo to dual-echo T_2^* estimates was previously reported by Beissner et al. (2010) using a finger tapping task; suggesting this issue is not specific to our dataset.

The 1E-T1C pipeline significantly outperformed the 1E pipeline, yet, it still yielded significantly less activation magnitude and extent than ME-ICA. Moreover, the 1E-T1C pipeline requires accurate logging of the interval of time that elapses between consecutive images and relies on obtaining accurate estimations of T_1 across the brain, which can be problematic when attempting imaging of certain structures such as the spinal cord (Xie et al., 2012). Here we demonstrate how ME-ICA can be a viable alternative for denoising cardiac-gated datasets that lacks these limitations. In addition, ME-ICA, in this context, provides a qualitative T_1 map as a major component. This T_1 map may be used for image registration or other analyses.

One important factor contributing to the superior performance of ME-ICA in cardiac-gated data is that ME-ICA was able to reliably detect components strongly correlated with T_1 artifacts. This is clearly exemplified by the fact that ME-ICA always assigned the highest rho (i.e., non-BOLD likelihood) and lowest kappa (i.e., BOLD likelihood) to a component whose time series strongly correlated (r=0.84 ± 0.16) with estimations of T_1-related signal shifts associated with the irregular acquisition intervals (Figure 7). Moreover, the average rho for this "cardiac-gated artifactual component" detected by ME-ICA was on average 237.1 ± 59.6. The average rho for all other noise components across all subjects in the non-gated datasets (which lacks such T_1 artifacts) was 28.7 ± 11.0, approximately an order of magnitude less. This shows how robustly ME-ICA can identify this T_1 artifactual component specific to gated datasets. It also highlights the potential of ME-ICA to identify and remove other T_1-related artifacts such as those associated with inflow effects in constant TR acquisitions, which are otherwise difficult to model and account for.

ME-ICA for constant-TR acquisitions

For non-gated data, the improvements derived from the use of ME-ICA, although also statistically significant relative to the other pipelines, were smaller in magnitude (see Figure 6). This may have been because all pre-processing pipelines included corrective steps for common artifacts such as slow signal drifts (via regression of Legendre polynomials), head motion (via rigid spatial realignment and regression of motion estimates and their first derivatives) and thermal noise (by

means of smoothing). Our data suggest that although ME-ICA statistically outperformed the other pipelines, the standard pre-processing pipeline was able to account, to a large extent, for the detrimental effects of these important contaminating sources, leaving a narrower margin of improvement for the ME-ICA algorithm. Nevertheless, the combined results from the gated and non-gated experiments highlight the versatility of ME-ICA to detect different types of artifacts without the need to adapt pre-processing pipelines to the specific characteristics of each dataset.

Detailed exploration of ME-ICA components marked as noise suggests that additional gains in sensitivity (relative to the other pipelines) came from the removal of physiological noise, scanner instabilities other than slow signal drifts, and residual head motion related artifacts. For all subjects and experimental paradigms, ME-ICA removed components with temporal and spatial patterns typical of these artifactual sources (Supplementary Figure 2.A – D). Yet, it is worth noticing that we still found a significant relationship between original per-run TSNR and detection percent of individual trials after ME-ICA (Figure 11.D). This suggests that residual traces of noise persist in the data despite the application of ME-ICA, and sensitivity is still dependent on original data quality. Additional improvements to the ME-ICA algorithm and optimization of multi-echo acquisition parameters may help mitigate this situation (please see Future Directions below for a more detailed discussion).

All improvements in activation extent and magnitude happened despite a significant decrease in estimated effect size from the 1E relative to the OC and ME-ICA pipelines. Although an additional significant decrease in effect size from OC to ME-ICA was observed, it was limited to only three tasks in the event related experiments. This observation, combined with the fact that OC is an intermediate component of the ME-ICA algorithm, suggests that the main cause of the decrease in effect size estimates is this OC step. This is in agreement with the simulations of BOLD contrast and OC weights shown in Figure 2. The OC scheme used here corresponds to a previously proposed voxel-wise linear combination of multi-echo time series designed to optimize BOLD contrast-to-noise (Posse et al., 1999); not simply BOLD contrast. With the exception of the cardiac-gated case, the OC pipeline always resulted in an increase in activation magnitude and a decrease in effect size relative to the 1E pipeline; confirming that the OC approach yielded the expected overall increase in BOLD contrast-to-noise; despite a concomitant decrease in effect size.

Overall experimental decreases in effect size from 1E to OC (Non-gated Block Data: 13.14%; Event-related: 8.86%) were larger than those in the simulations (6.63%). It is possible that small errors in T_2^* estimates, given the limited number of echoes available, or the fact that we did not correct for the use of parallel imaging (Poser et al., 2006), may have affected weight computations and produced the observed additional decrease in effect size. Also, other weighting methods, such as those based on temporal signal-to-noise estimates (Poser et al., 2006), may help alleviate

these issues. Nevertheless, despite the decrease in effect size estimates, the OC step led to an overall improvement in BOLD contrast-to-noise as evidenced by the increases in activation extent and activation magnitude. The same is true for ME-ICA.

Finally, it is worth noting that the performance of ME-ICA was tested for the two most common task-based experimental paradigms (block and event related). Moreover, for event related paradigms, we used five tasks expected to evoke activity in many different brain regions, including but not limited to, primary sensory motor cortex. In all instances, ME-ICA outperformed the other pipelines in terms of activation extent and magnitude. In terms of detectability of individual events, ME-ICA did better in all tasks, although pairwise comparison only reached significance for three of the five tasks. These results suggest that our conclusions regarding the better performance of ME-ICA over traditional single-echo fMRI and optimal combination of multi-echo data generalize well across tasks and paradigms. Prior studies have demonstrated the ability of ME-ICA to improve functional connectivity analysis, especially for subcortical regions (Kundu et al., 2013). Our results suggest that ME-ICA is also a viable denoising option for traditional task-based studies. Further research should evaluate how ME-ICA may help increase the sensitivity and interpretability of studies that use naturalistic stimuli for purposes such as decoding (Nishimoto et al., 2011) or looking for patterns of inter-subject correlation (Hasson et al., 2004).

Limitations of the current study

In this study, we compared ME-ICA to four other pre-processing pipelines, of which only one mimics single-echo fMRI protocols. Although most common pre-processing steps were included in the 1E pipeline, more aggressive denoising could be accomplished in single-echo datasets using additional steps. For example, different aspects of physiological noise can be corrected if concurrent physiological traces (e.g., cardiac and respiration) are available (Birn et al., 2008; Chang et al., 2009; Glover et al., 2000); or even in their absence (Beall and Lowe, 2007; Behzadi et al., 2007). Manual and automatic single-echo ICA-based denoising procedures (Pruim et al., 2015; Salimi-Khorshidi et al., 2014) can also help remove additional traces of noise. Prior research has shown that there can be substantial inter-subject differences in terms of optimal pre-processing pipelines (Strother et al., 2004). It is possible that comparison of ME-ICA against these other single-echo pre-processing pipelines, including subject-specific ones, would show relatively less improvements.

Also, the present study focused only on different aspects of sensitivity. The performance of ME-ICA ought to also be evaluated in terms of test-retest reproducibility across scans, subjects and sites. For example, more accurate accounting of hardware instabilities, where ME-ICA tends to do well, may help improve the reproducibility of single-subject results across sites, an important standing challenge for future clinical fMRI applications. Further research should

evaluate these other scenarios before making any categorical claims about the superiority of multi-echo denoising approaches over single-echo.

Finally, in this study we used the middle echo as a proxy for standard single-echo fMRI. In this manner, we were able to compare pre-processing pipelines for single and multi-echo datasets using the same data (with the same artifacts) for all pipelines. Yet, the acquisition of additional echoes comes at a cost in temporal resolution. For example, single-echo runs with the same spatial resolution, in-plane acceleration, and number of slices of the block-design dataset could have been acquired using a TR of 1875ms, instead of the current one of 2500ms, if the echo time were that of middle echo (TE=31.7ms). It is possible that additional data points, derived from shorter TRs, could help increase the statistical power of equivalent "true" single-echo datasets, and reduce the differences presented here.

Nevertheless, despite the limitations cited above, our results agree with prior reports of the better performance of ME-ICA for functional connectivity studies (Kundu et al., 2012). These, combined with experimental demonstration of its ability to separate BOLD-like (as in ultra-slow block designs) from non-BOLD (scanner related) slow signal fluctuations (Evans et al., 2015), and its ability to remove simultaneous multi-slice related artifacts (Olafsson et al., 2015), suggest that multi-echo fMRI acquisition strategies combined with ME-ICA are a versatile and powerful alternative to current single-echo fMRI acquisition and pre-processing schemes.

Future directions for ME-ICA

Although ME-ICA was able to remove greater amounts of noise in the data than the other pipelines were, three different findings suggest the ME-ICA procedure could still be further improved. First, for the ME-ICA per-trial analyses, only 46% of individual trials were detected in the fMRI maps. This, despite strong evidence of subject compliance based on eye tracker and button box responses. Second, we found a significant relationship between data quality (i.e., TSNR) and individual trial percent detection after ME-ICA. These two observations suggest that there is room both for improvements in sensitivity and for the removal of residual noise. Third, detailed evaluation of ME-ICA outputs revealed consistent misclassification of a few clear noise-like components (see examples in Figure 12) as "not-noise" in all subjects. This third observation confirms that the ME-ICA algorithm did not correctly eliminate all potential noise components. In the remainder of this section we discuss several ways in which the ME-ICA algorithm could be improved, as well as the need for additional systematic evaluations to optimize multi-echo acquisitions for ME-ICA.

The following are the main components of the ME-ICA algorithm (see (Kundu et al., 2012) for a detailed description): (1) generation of OC time series (i.e., a T_2^* weighted voxel-wise linear combination of all echoes); (2) spatial ICA over the OC time series to determine spatially independent signal components in the data; (3) computation of a per-component feature set aimed primarily at characterizing its

TE-dependence profile; (4) classification of components as noise/not-noise running an empirically defined decision tree over the feature set; and (5) removal of noise components from the data. Of these, components 2, 3 and 4 are the best targets for algorithmic tuning.

Spatial ICA, the most common form of ICA in fMRI, is designed to separate spatially independent sources for which only linear mixtures of the sources themselves (e.g., voxel-wise time series) are available (Mckeown et al., 1998). When applied to fMRI data, spatial ICA has proven successful at reliably extracting both noise (Thomas et al., 2002) and biologically meaningful (Smith et al., 2009) signal components. While many ME-ICA components show clear noise-like (their rho is among the highest and their kappa among the lowest; Figure 12.B) and BOLD-like (their kappa is among the highest and their rho among the lowest; Figure 12.D) profiles, there is also a non-negligible number of components whose kappa and rho suggest they constitute a mixture of both BOLD and non-BOLD effects (their position in the kappa-rho plot is near the 45° degree line; Figure 12.A, dashed black line). Keeping components with a mixture of TE and non-TE dependence may translate into suboptimal denoising, as non-TE components remain in the data after the ME-ICA step. Removing them risks eliminating BOLD-effects of neuronal origin (e.g., false negatives). It may be possible that alternative source separation methods could better extract signal components with purer TE dependence profiles. Potential candidates include regionally restricted spatial ICA (e.g., a searchlight-like approach), independent vector analysis (Adali et al., 2014), temporal ICA (Smith et al., 2012) and deep learning approaches (Plis et al., 2014). All these alternatives have been successfully applied to fMRI datasets in contexts other than denoising. Future research should evaluate their performance for the specific purpose of multi-echo based denoising.

Kappa and rho are ME-ICA's primary way to characterize the overall TE-dependence profiles of ICA components. Kappa is a weighted grand-average of how well the magnitude of the component varies linearly with TE across the imaged volume of brain (as expected for pure BOLD components). Its counterpart, rho, tells on average how much the component's magnitude remains constant across echoes. In theory, noise components (e.g., scanning artifacts and head motion artifacts) should have high rho and negligible kappa, while BOLD-like components (e.g., task-induced activation and intrinsic resting-state fluctuations) should have high kappa and negligible rho. In practice, those expectations are not always met, with some noise components—easily identifiable as noise by their spatio-temporal patterns—having similar kappa and rho (e.g., sitting near the 45° line in Figure 12.A) or, even higher kappa than rho (Figure 12.C). ME-ICA deals with this reality by using additional features during the final classification step, including explained variance, relative percentage of voxels fitting the two TE dependence profiles of interest (e.g., no-dependence or linear dependence), and overlap between component maps and voxels that best fit each TE dependence profile. Moreover, ME-ICA does not operate in absolute terms of kappa and rho, but on their rank and relative position with respect to empirically observed inflection points (e.g., an elbow) in both kappa and

rho spectrums that tend to indicate the border between mostly BOLD and mostly noise components (see Figures 4 and 5 in (Kundu et al., 2012)). Despite all these safeguards, our data show how noise components, in some instances, are incorrectly labeled as BOLD-like (Figure 12.C), most likely because these components have similar kappa and rho values, but also because they are located near the above-mentioned empirically defined inflection points. We believe that refinements in the computation of kappa and rho (e.g., better accounting for outliers in the TE fits or using a different weighting scheme when averaging across the brain) may help improve the profiling of components as BOLD and non-BOLD. In addition, the ME-ICA algorithm may also benefit from the inclusion of additional, not necessarily TE-based, features that may help better characterize noise components in a manner similar to how other automatic ICA denoising algorithms do (Pruim et al., 2015; Salimi-Khorshidi et al., 2014).

Finally, additional improvements to ME-ICA could come from modifying or substituting its current empirically defined decision tree by data driven alternatives (e.g., logistic regression, clustering) that may better combine the rich information gathered by the current, or a potentially expanded, feature set.

In addition to potential updates to the ME-ICA algorithm itself, several questions regarding how to best optimize multi-echo data acquisition and pre-processing for ME-ICA still require systematic empirical evaluation. For example, most prior studies that use ME-ICA acquired three or four equally spaced echoes; partly to avoid incurring large losses in temporal resolution. Computation of kappa and rho, the main decision criteria, relies on voxel-wise linear fits based on as many points as there are available echoes. The same is true for computation of static voxel-wise T_2^* maps later used for generation of OC time series (Eq. 2). Although three points are sufficient to compute a linear fit (please see Supplementary Figure 3 for representative T_2^* maps and goodness of fit in terms of the coefficient of determination R^2), additional points can help improve the quality of the fits as long as these additional points are not excessively noisy. Multi-echo sequences can acquire more than three echoes, yet any additional echoes will have lower signal-to-noise ratio as the longer the time interval between the radio-frequency pulse and read-out window the lower the amount of signal available. Future research should evaluate the optimal number of echoes, and how much the specific echo times matter, for the purpose of ME-ICA denoising. These additional investigations should consider the use of multi-echo multiband sequences which now permit acquisition of additional echoes without incurring concomitant loses in temporal resolution. Several recent studies (Boyacioğlu et al., 2015; Olafsson et al., 2015) have reported that combined multi-echo/multiband approaches can help better account for high frequency artifacts, help with removal of physiological noise, and ultimately improve the spatial specificity and sensitivity of regular multi-echo fMRI. In fact, initial evaluation of ME-ICA with multi-echo/multiband datasets suggest that ME-ICA can clearly benefit from these richer datasets (Olafsson et al., 2015).

The effect of other important acquisition parameters with the potential to alter the noise profile of the data such as amount of in-plane acceleration, multi-slice acceleration factor or flip angle should also be evaluated systematically. Finally, another important factor to consider is the effect that other pre-processing steps may have on the ME-ICA algorithm. For example, any registration step will come at the cost of spatial smoothing. While spatial smoothing may help remove thermal noise and aid with ICA convergence, it may also produce undesired mixing of BOLD and non-BOLD signals, making their separation and consequent classification more difficult. It may be possible that ME-ICA may benefit from conducting the TE-dependence analysis at an earlier pre-processing step than as currently implemented here. Empirical optimization of all these factors, combined with improvements in the algorithm itself, should help further improve the denoising capabilities of the ME-ICA methodology, and perhaps help fMRI achieve the necessary single-subject levels of sensitivity, specificity and reproducibility to achieve its long-term goal of entering routine clinical practice.

CONCLUSION

In this study, we evaluated the performance of the ME-ICA denoising technique for task-based fMRI studies at the single-subject level under different experimental scenarios. In all instances, ME-ICA showed superior sensitivity to the other two alternatives under evaluation, suggesting its potential suitability for clinical applications where group averaging is not possible. ME-ICA performed especially well in cardiac-gated datasets, where we demonstrated how it was able to reliably remove T_1 artifacts associated with irregular repetition times.

ACKNOWLEDGEMENTS

This research was possible thanks to the support of the National Institute of Mental Health Intramural Research Program. Portions of this study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (*biowulf.nih.gov*). This study is part of NIH clinical protocol number NCT00001360, protocol ID 93-M-0170 and annual report ZIAMH002783-14. Dr. Caballero-Gaudes was supported by the Spanish Ministry of Economy and Competitiveness, through grant PSI 2013-42343 Neuroimagen Multimodal and the Severo Ochoa Programme for Centres/Units of Excellence in R&D (SEV-2015-490).

REFERENCES

Adali, T., Anderson, M., Fu, G., 2014. Diversity in independent component and vector analyses. IEEE Signal Process. Mag. 31, 18–33. doi:10.1109/MSP.2014.2300511

1059 Beall, E.B., Lowe, M.J., 2007. Isolating physiologic noise sources with independently
1060 determined spatial measures. Neuroimage 37, 1286–1300.
1061 doi:10.1016/j.neuroimage.2007.07.004
1062 Behzadi, Y., Restom, K., Liau, J., Liu, T.T., 2007. A component based noise correction
1063 method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37, 90–
1064 101. doi:10.1016/j.neuroimage.2007.04.042
1065 Beissner, F., Baudrexel, S., Volz, S., Deichmann, R., 2010. Dual-echo EPI for non-
1066 equilibrium fMRI - Implications of different echo combinations and masking
1067 procedures. Neuroimage 52, 524–531. doi:10.1016/j.neuroimage.2010.04.243
1068 Beissner, F., Deichmann, R., Baudrexel, S., 2011. FMRI of the brainstem using dual-
1069 echo EPI. Neuroimage 55, 1593–1599. doi:10.1016/j.neuroimage.2011.01.042
1070 Bianciardi, M., Fukunaga, M., van Gelderen, P., Horovitz, S.G., de Zwart, J.A., Shmueli,
1071 K., Duyn, J.H., 2009. Sources of functional magnetic resonance imaging signal
1072 fluctuations in the human brain at rest: a 7 T study. Magn. Reson. Imaging 27,
1073 1019–29. doi:10.1016/j.mri.2009.02.004
1074 Birn, R.M., Smith, M.A., Jones, T.B., Bandettini, P.A., 2008. The respiration response
1075 function: the temporal dynamics of fMRI signal fluctuations related to changes
1076 in respiration. Neuroimage 40, 644–54. doi:10.1016/j.neuroimage.2007.11.059
1077 Boyacioğlu, R., Schulz, J., Koopmans, P.J., Barth, M., Norris, D.G., 2015. Improved
1078 sensitivity and specificity for resting state and task fMRI with multiband multi-
1079 echo EPI compared to multi-echo EPI at 7T. Neuroimage 119, 352–361.
1080 doi:10.1016/j.neuroimage.2015.06.089
1081 Bright, M.G., Murphy, K., 2013. Removing motion and physiological artifacts from
1082 intrinsic BOLD fluctuations using short echo data. Neuroimage 64, 526–537.
1083 doi:10.1016/j.neuroimage.2012.09.043
1084 Buur, P.F., Poser, B.A., Norris, D.G., 2009. A dual echo approach to removing motion
1085 artefacts in fMRI time series. NMR Biomed. 22, 551–560.
1086 doi:10.1002/nbm.1371
1087 Chang, C., Cunningham, J.P., Glover, G.H., 2009. Influence of heart rate on the BOLD
1088 signal: the cardiac response function. Neuroimage 44, 857–69.
1089 doi:10.1016/j.neuroimage.2008.09.029
1090 Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic
1091 resonance neuroimages. Comput. Biomed. Res. 29, 162–73.
1092 Evans, J.W., Kundu, P., Horovitz, S.G., Bandettini, P. a., 2015. Separating slow BOLD
1093 from non-BOLD baseline drifts using multi-echo fMRI. Neuroimage 105, 189–
1094 197. doi:10.1016/j.neuroimage.2014.10.051
1095 Feinberg, D.A., Moeller, S., Smith, S.M., Auerbach, E., Ramanna, S., Glasser, M.F.,
1096 Miller, K.L., Ugurbil, K., Yacoub, E., 2010. Multiplexed Echo Planar Imaging for
1097 Sub-Second Whole Brain FMRI and Fast Diffusion Imaging. PLoS One 5, e15710.
1098 doi:10.1371/journal.pone.0015710
1099 Geyer, S., Ledberg, a, Schleicher, a, Kinomura, S., Schormann, T., Bürgel, U.,
1100 Klingberg, T., Larsson, J., Zilles, K., Roland, P.E., 1996. Two different areas within
1101 the primary motor cortex of man. Nature. doi:10.1038/382805a0
1102 Glover, G.H., Li, T.Q., Ress, D., 2000. Image-based method for retrospective
1103 correction of physiological motion effects in fMRI: RETROICOR. Magn. Reson.
1104 Med. 44, 162–7.

Gonzalez-Castillo, J., Duthie, K.N., Saad, Z.S., Chu, C., Bandettini, P. a, Luh, W.-M., 2012a. Effects of image contrast on functional MRI image registration. Neuroimage 67, 163–174. doi:10.1016/j.neuroimage.2012.10.076

Gonzalez-Castillo, J., Saad, Z.S., Handwerker, D.A., Inati, S.J., Brenowitz, N., Bandettini, P.A., 2012b. Whole-brain, time-locked activation with simple tasks revealed using massive averaging and model-free analysis. Proc. Natl. Acad. Sci. U. S. A. 109, 5487–92. doi:10.1073/pnas.1121049109

Gowland, P.A., Bowtell, R., 2007. Theoretical optimization of multi-echo fMRI data acquisition. Phys. Med. Biol. 52, 1801–1813. doi:10.1088/0031-9155/52/7/003

Greve, D.N., Brown, G.G., Mueller, B.A., Glover, G., Liu, T.T., 2013. A survey of the sources of noise in fMRI. Psychometrika 78, 396–416. doi:10.1007/s11336-012-9294-0

Guimares, a R., Melcher, J.R., Talavage, T.M., Baker, J.R., Ledden, P., Rosen, B.R., Kiang, N.Y.S., Fullerton, B.C., Weisskoff, R.M., 1998. Imaging subcortical activity in humans. Hum Brain Map 6, 33–41.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject Synchronization of Cortical Activity During Natural Vision. Science (80-.). 303, 1634–1640.

Kundu, P., Brenowitz, N.D., Voon, V., Worbe, Y., Vértes, P.E., Inati, S.J., Saad, Z.S., Bandettini, P. a, Bullmore, E.T., 2013. Integrated strategy for improving functional connectivity mapping using multiecho fMRI. Proc. Natl. Acad. Sci. U. S. A. 110, 16187–92. doi:10.1073/pnas.1301725110

Kundu, P., Inati, S.J., Evans, J.W., Luh, W.-M., Bandettini, P. a, 2012. Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. Neuroimage 60, 1759–70. doi:10.1016/j.neuroimage.2011.12.028

Kundu, P., Inati, S.J., Evans, J.W., Luh, W.-M., Bandettini, P. a., 2014. Differentiating BOLD and non-BOLD signals in fMRI time series from anesthesized rats using multi-echo EPI at 11.7T. Neuroimage 60, 1759–1770. doi:10.1016/j.neuroimage.2011.12.028

Lombardo, M. V, Auyeung, B., Holt, R., Waldman, J., Ruigrok, A., Mooney, N., Bullmore, E.T., Baron-Cohen, S., Kundu, P., 2015. Improving effect size and power with multi-echo fMRI and its impact on understanding the neural systems supporting mentalizing. bioRxiv.

Malikovic, A., Amunts, K., Schleicher, A., Mohlberg, H., Eickhoff, S.B., Wilms, M., Palomero-Gallagher, N., Armstrong, E., Zilles, K., 2006. Cytoarchitectonic Analysis of the Human Extrastriate Cortex in the Region of V5/MT+: A Probabilistic, Stereotaxic Map of Area hOc5. Cereb. Cortex 17, 562–574. doi:10.1093/cercor/bhj181

Mckeown, M.J., Makeig, S., Brown, G.G., Jung, T., Kindermann, S.S., Bell, A.J., Sejnowski, T.J., 1998. Analysis of fMRI Data by Blind Separation Into Independent Spatial Components. Hum. Brain Mapp. 188, 160–188.

Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., Gallant, J.L., 2011. Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. Curr. Biol. 21, 1641–1646. doi:10.1016/j.cub.2011.08.031

Olafsson, V., Kundu, P., Wong, E.C., Bandettini, P. a., Liu, T.T., 2015. Enhanced

identification of BOLD-like components with multi-echo simultaneous multi-slice (MESMS) fMRI and multi-echo ICA. Neuroimage 112, 43–51. doi:10.1016/j.neuroimage.2015.02.052

Parrish, T.B., Gitelman, D.R., LaBar, K.S., Mesulam, M.-M., 2000. Impact of signal-to-noise on functional MRI. Magn. Reson. Med. 44, 925–932. doi:10.1002/1522-2594(200012)44:6<925::AID-MRM14>3.0.CO;2-M

Parsons, C.E., Young, K.S., Joensson, M., Brattico, E., Hyam, J.A., Stein, A., Green, A.L., Aziz, T.Z., Kringelbach, M.L., 2014. Ready for action: a role for the human midbrain in responding to infant vocalizations. Soc. Cogn. Affect. Neurosci. 9, 977–984. doi:10.1093/scan/nst076

Peirce, J.W., 2008. Generating stimuli for neuroscience using PsychoPy. Front. Neuroinform. 2, 1–8. doi:10.3389/neuro.11.010.2008

Plis, S.M., Hjelm, D.R., Slakhutdinov, R., Allen, E. a., Bockholt, H.J., Long, J.D., Johnson, H., Paulsen, J., Turner, J., Calhoun, V.D., 2014. Deep learning for neuroimaging: A validation study. Front. Neurosci. 8, 1–11. doi:10.3389/fnins.2014.00229

Poser, B. a., Norris, D.G., 2009. Investigating the benefits of multi-echo EPI for fMRI at 7T. Neuroimage 45, 1162–1172. doi:10.1016/j.neuroimage.2009.01.007

Poser, B. a., Versluis, M.J., Hoogduin, J.M., Norris, D.G., 2006. BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. Magn. Reson. Med. 55, 1227–1235. doi:10.1002/mrm.20900

Posse, S., 2012. Multi-echo acquisition. Neuroimage 62, 665–671. doi:10.1016/j.neuroimage.2011.10.057

Posse, S., Wiese, S., Gembris, D., Mathiak, K., Kessler, C., Grosse-Ruyken, M.L., Elghahwagi, B., Richards, T., Dager, S.R., Kiselev, V.G., 1999. Enhancement of BOLD-contrast sensitivity by single-shot multi-echo functional MR imaging. Magn. Reson. Med. 42, 87–97. doi:10.1002/(SICI)1522-2594(199907)42:1<87::AID-MRM13>3.0.CO;2-O

Pruim, R.H.R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J.K., Beckmann, C.F., 2015. ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. Neuroimage 112, 267–277. doi:10.1016/j.neuroimage.2015.02.064

Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., Smith, S.M., 2014. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90, 449–468. doi:10.1016/j.neuroimage.2013.11.046

Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., 2009. Correspondence of the brain' s functional architecture during activation and rest. Proc. Natl. Acad. Sci. U. S. A. 106, 13040–13045.

Smith, S.M., Miller, K.L., Moeller, S., Xu, J., Auerbach, E.J., Woolrich, M.W., Beckmann, C.F., Jenkinson, M., Andersson, J., Glasser, M.F., Van Essen, D.C., Feinberg, D.A., Yacoub, E.S., Ugurbil, K., 2012. Temporally-independent functional modes of spontaneous brain activity. Proc. Natl. Acad. Sci. U. S. A. 109, 3131–6. doi:10.1073/pnas.1121329109

Speck, O., Hennig, J., 1998. Functional imaging by I0- and T2*-parameter mapping

using multi-image EPI. Magn. Reson. Med. 40, 243–8. doi:10.1002/mrm.1910400210

Strother, S., La Conte, S., Kai Hansen, L., Anderson, J., Zhang, J., Pulapura, S., Rottenberg, D., 2004. Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. A preliminary group analysis. Neuroimage 23, 196–207. doi:10.1016/j.neuroimage.2004.07.022

Tanabe, J., Miller, D., Tregellas, J., Freedman, R., Meyer, F.G., 2002. Comparison of detrending methods for optimal fMRI preprocessing. Neuroimage 15, 902–907. doi:10.1006/nimg.2002.1053

Thomas, C.G., Marshman, R.A., Menon, R.S., 2002. Noise Reduction in BOLD-Based fMRI Using Component Analysis. Neuroimage 17, 1521–1537. doi:10.1006/nimg.2002.1200

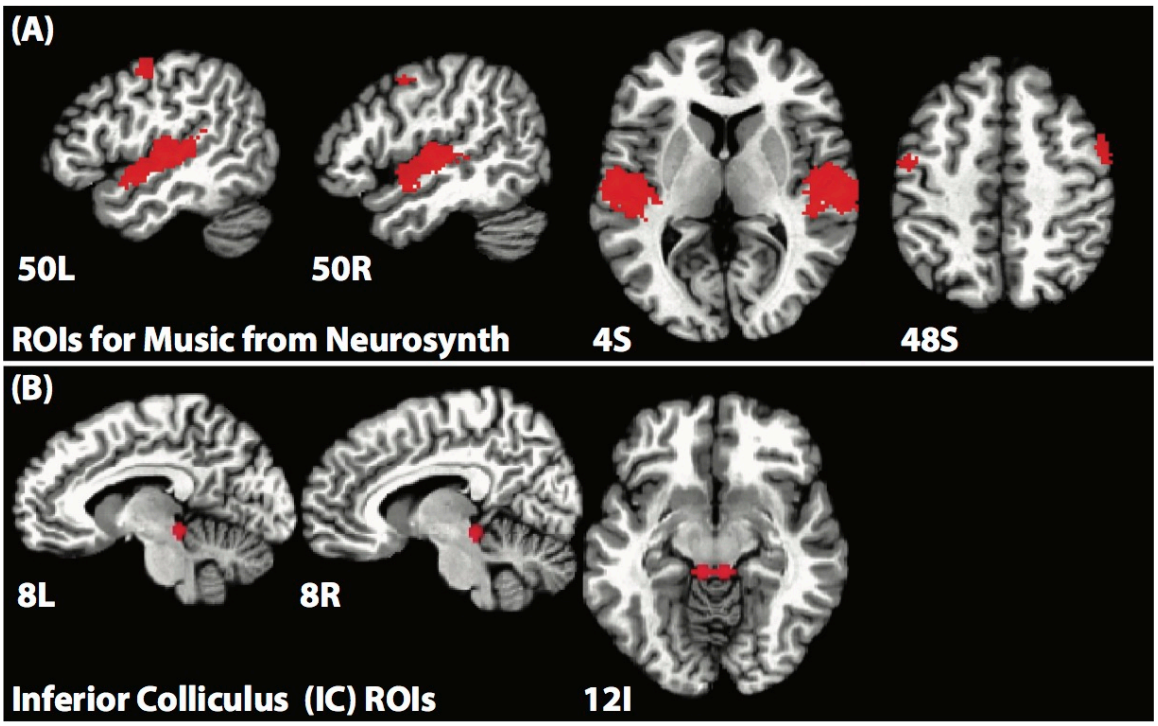
Vosskühler, A., Nordmeier, V., Kuchinke, L., Jacobs, A.M., 2008. OGAMA (Open Gaze and Mouse Analyzer): open-source software designed to analyze eye and mouse movements in slideshow study designs. Behav. Res. Methods 40, 1150–1162. doi:10.3758/BRM.40.4.1150

Wansapura, J.P., Holland, S.K., Dunn, R.S., Ball, W.S., 1999. NMR relaxation times in the human brain at 3.0 tesla. J. Magn. Reson. Imaging 9, 531–8.

Xie, G., Piché, M., Khoshnejad, M., Perlbarg, V., Chen, J.I., Hoge, R.D., Benali, H., Rossignol, S., Rainville, P., Cohen-Adad, J., 2012. Reduction of physiological noise with independent component analysis improves the detection of nociceptive responses with fMRI of the human spinal cord. Neuroimage 63, 245–252. doi:10.1016/j.neuroimage.2012.06.057

Yarkoni, T., Poldrack, R. a, Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8, 665–70. doi:10.1038/nmeth.1635

FIGURE LEGENDS



1. Target ROIs for the performance analysis of block-design dataset. (A) Target ROI obtained with the *Neurosynth* tool for the concept "music". (B) Spherical ROIs sitting on bilateral inferior colliculus.

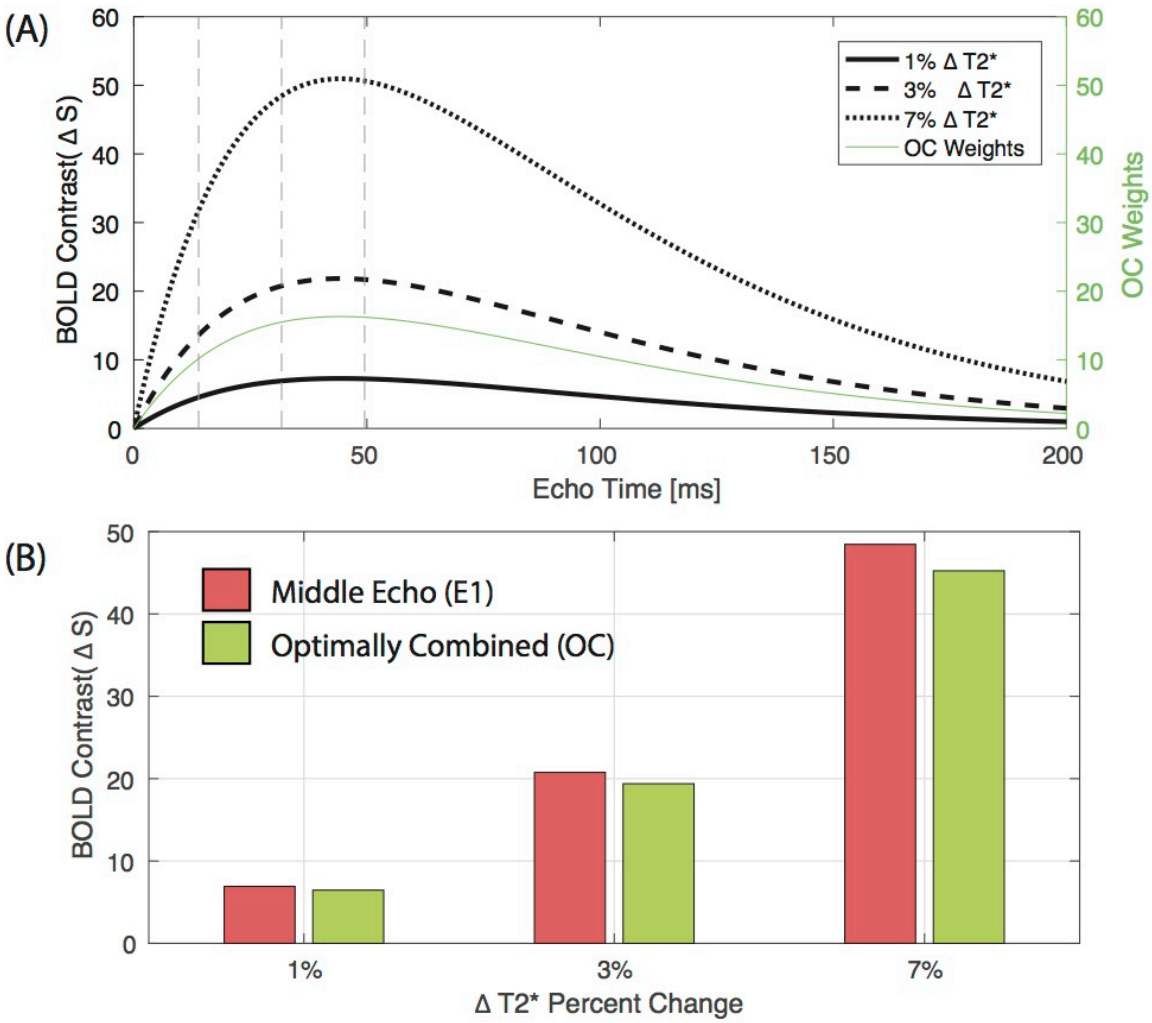


Figure 2. (A) Simulated BOLD contrast curves for echo times ranging from 0 to 200 ms for three different amounts of task-induced ΔT2* (i.e., 1%, 3% and 7%; black curves). Estimated OC weights according to Eq. 1 for the same range of echo times are shown in green. (b) Simulated effect size (i.e., BOLD contrast) for the 1E and OC pipelines for the three different ΔT2* scenarios.

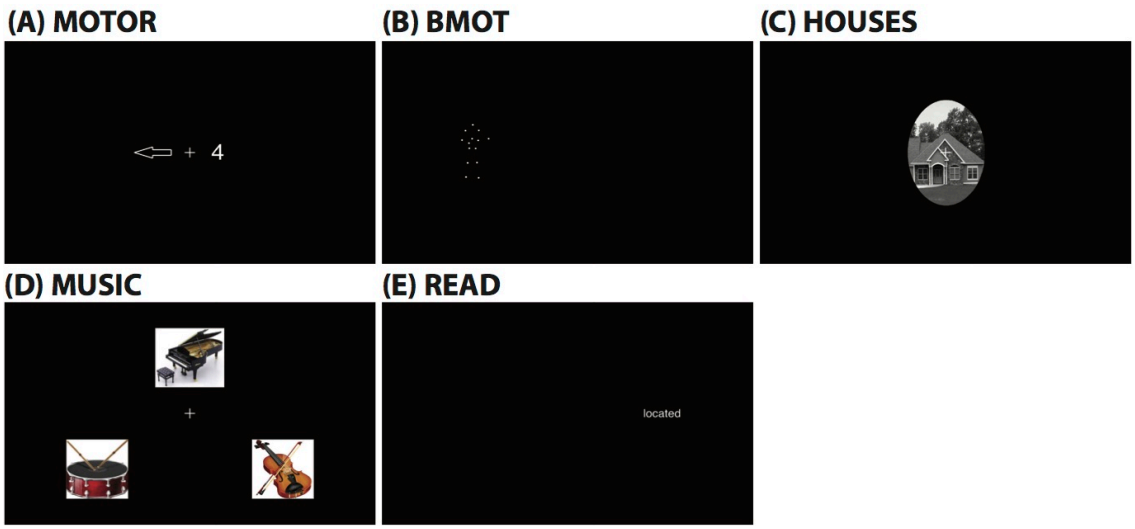


Figure 3. Sample screenshots for each of the five tasks under evaluation in the rapid event-related experiments. (A) Motor task. (B) Biological motion detection task. (C) Houses visualization task. (D) Music listening task. (E) Sentence reading task.

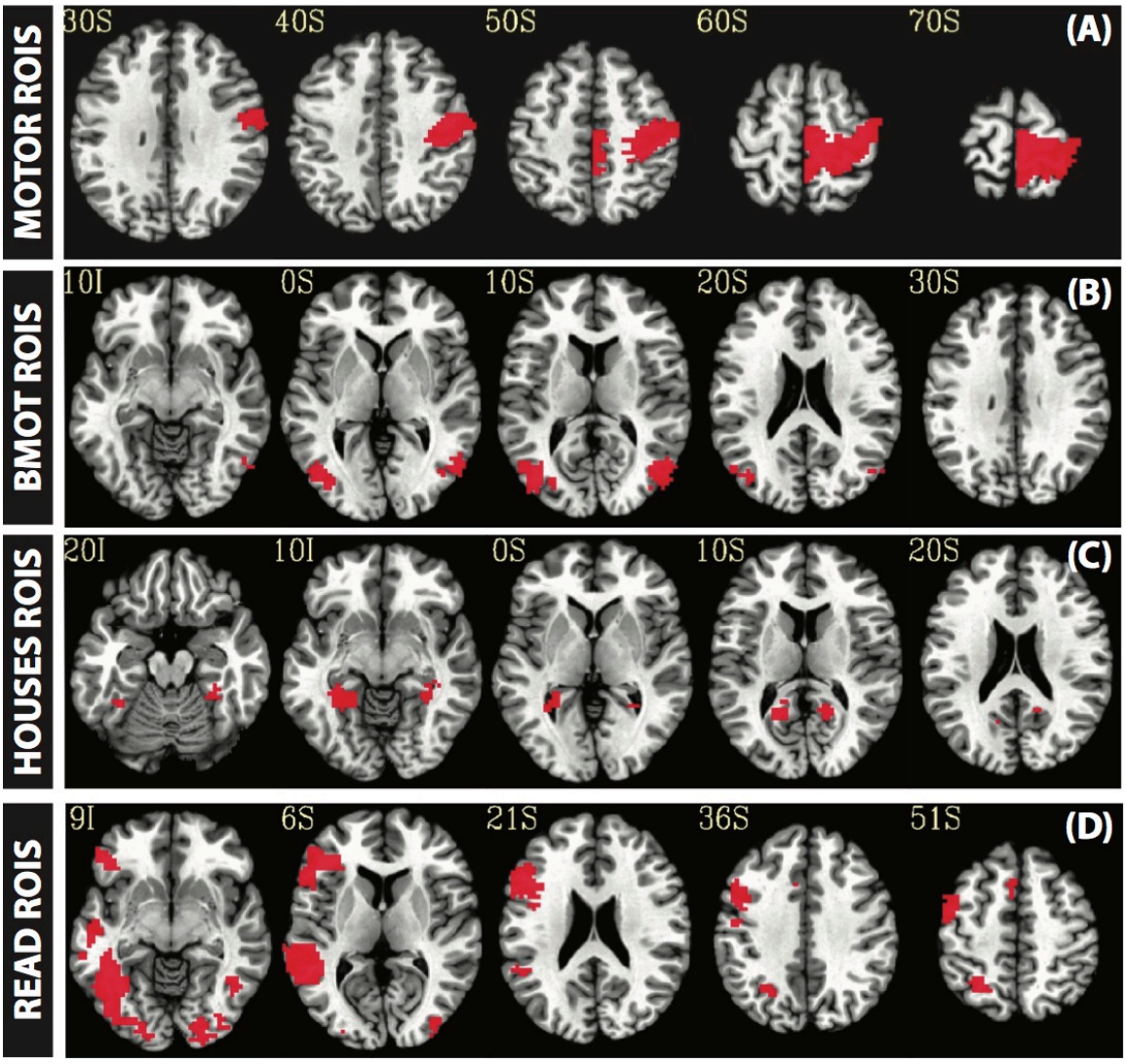


Figure 4. Target ROI sets for the rapid event-related experiments. (A) Target ROIs for the motor task. (B) Target ROIs for the biological motion observation task. (C) Target ROIs for the house visualization task. (D) Target ROIs for the sentence reading task.

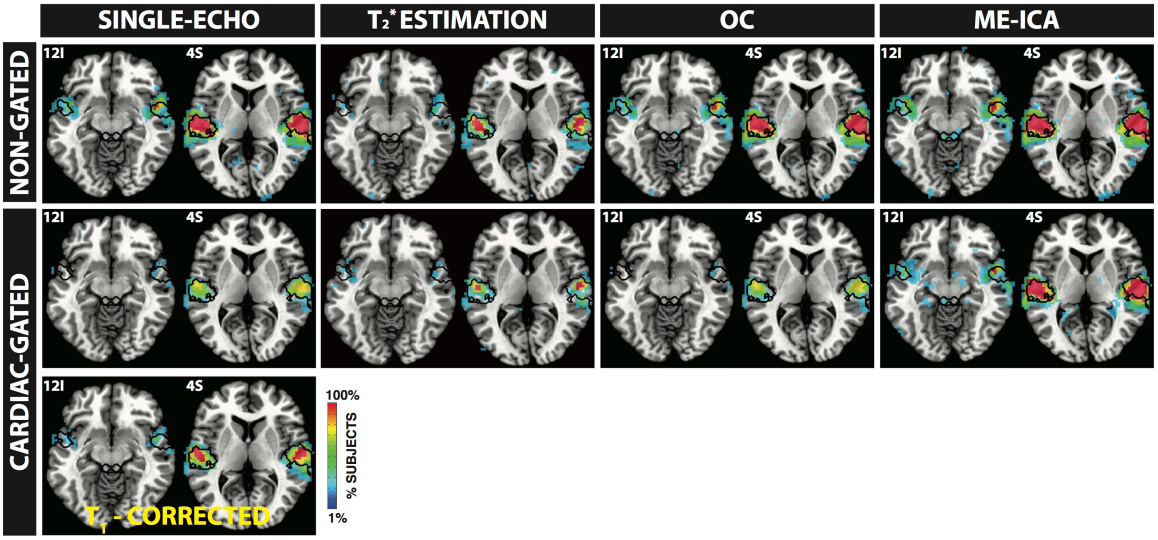


Figure 5. Probabilistic maps of activation across subjects for all pre-processing pipelines for both acquisition schemes (cardiac-gated and non-gated). The color of a voxel indicates the percentage of subjects for which that voxel became significant at $p_{FDR}<0.05$. In all instances, activity was detected in portions of the superior temporal gyrus. IC activation was present more consistently across subjects only for the ME-ICA pipeline (see Table 1). Target ROIs are shown with a black contour.

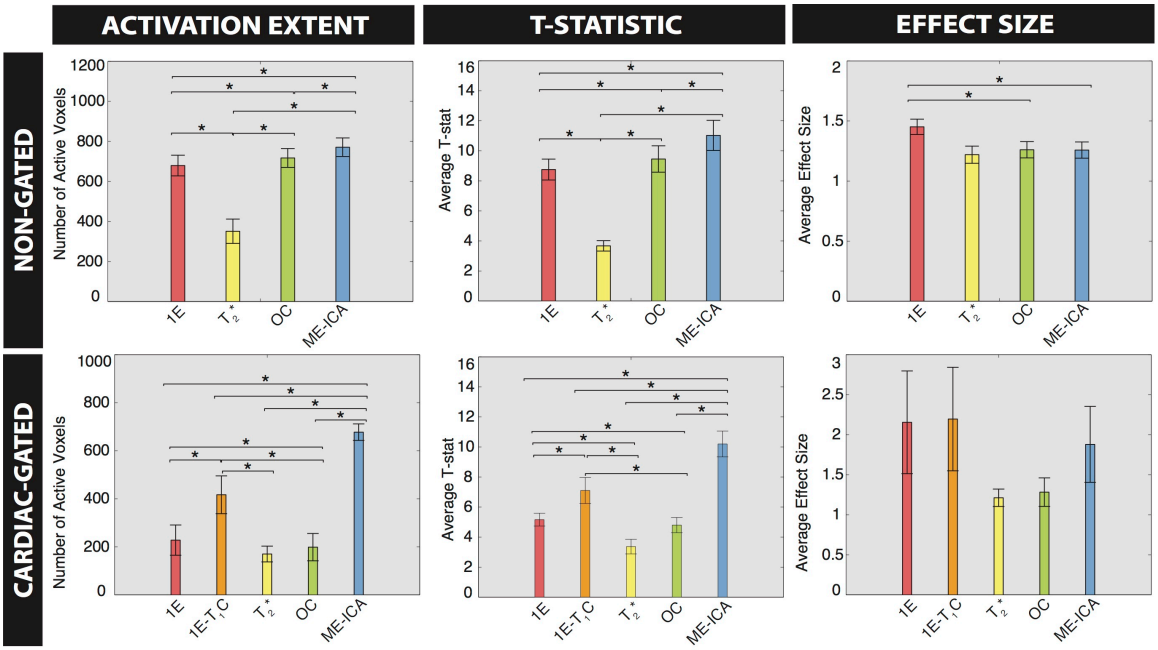


Figure 6. Performance analysis results for the block-design dataset. The top row shows the results for the non-gated data and the bottom row for the cardiac-gated. Post-hoc paired T-tests that reached significance ($p<0.05$) are marked with an asterisk.

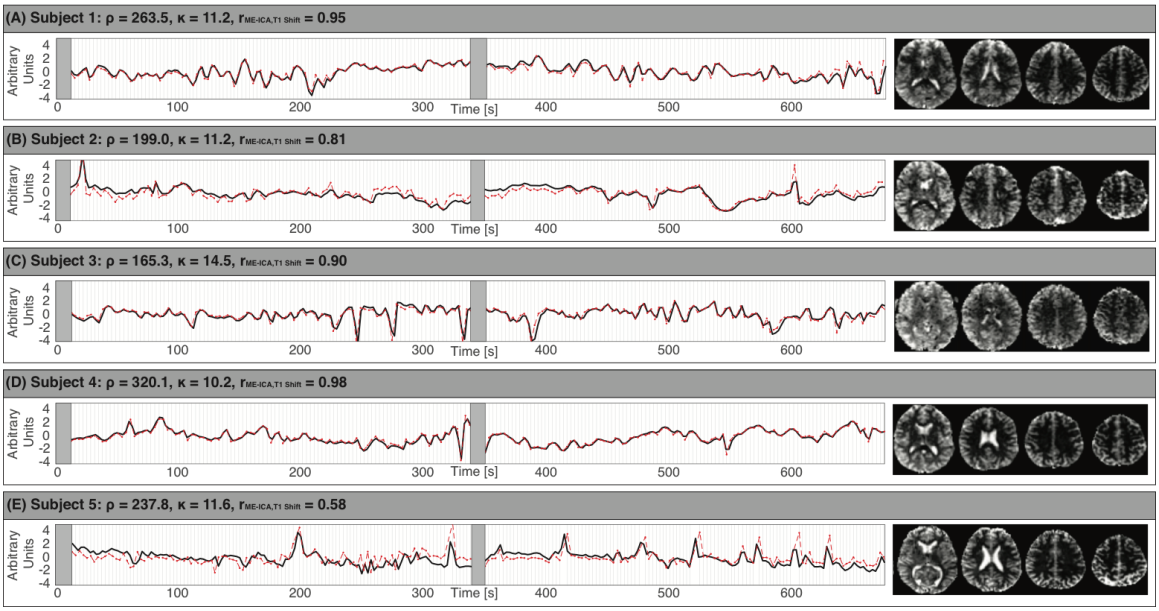


Figure 7. ME-ICA components marked as noise and showing the largest correlation with the estimation of T_1-related signal changes derived from irregular TR in cardiac-gated acquisitions. For each subject we show the ME-ICA component spatial map on the right, its associated time series in black and the estimation of T_1-related signal change for that subject in dashed red. All time series are normalized (mean of zero and standard deviation of one) so that similarities in shape can be better visualized. Light grey vertical lines indicate actual acquisition times of the MRI volumes. Dark grey rectangles signal the times associated with discarded volumes at the beginning of each run.

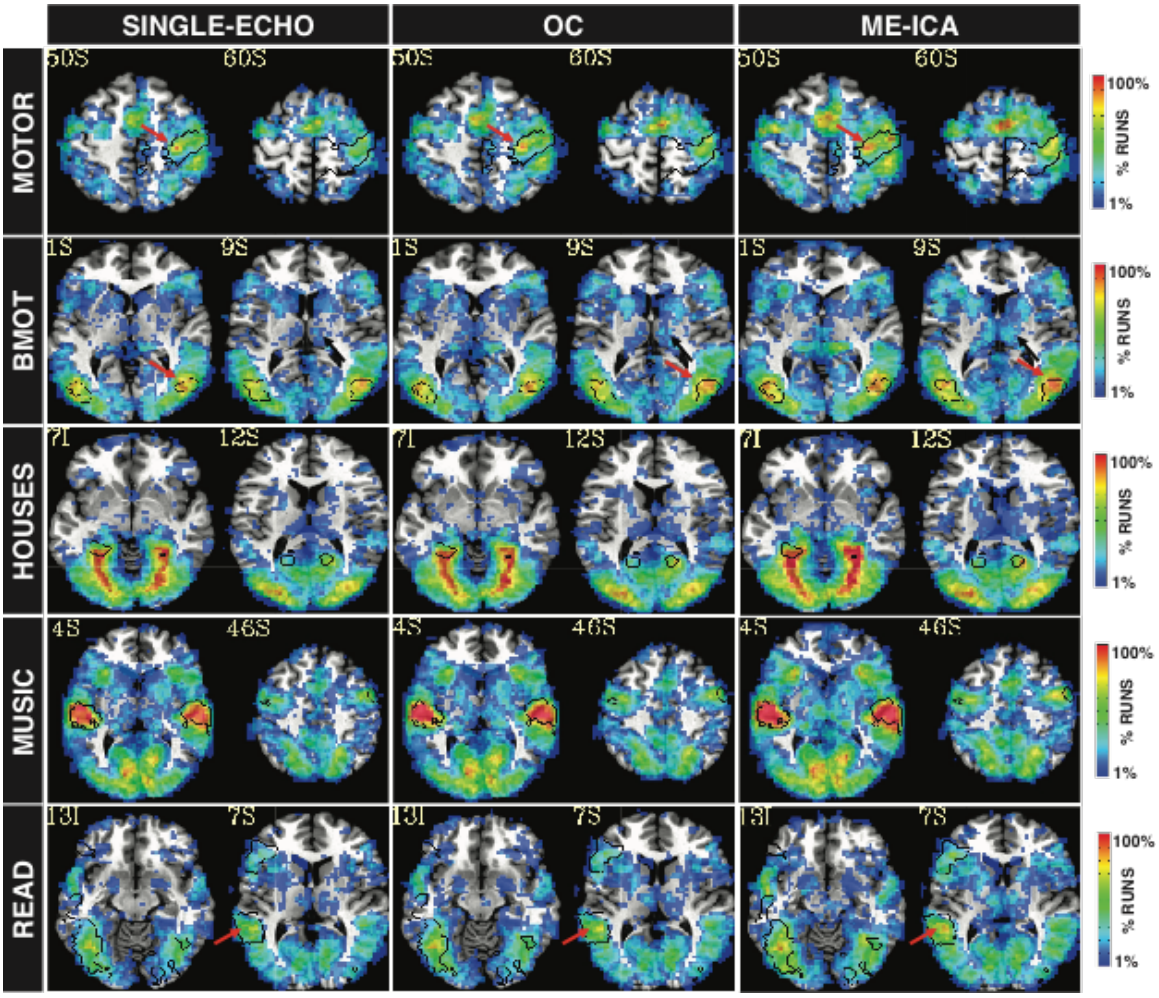


Figure 8. Probabilistic maps of activation across runs for the per-task (rapid event-related) analysis for all five tasks and three pre-processing pipelines. While a total of 16 runs contributed to the probabilistic maps for the BMOT, HOUSES, MUSIC and READ tasks, only 13 did for the MOTOR task. These 13 activation maps correspond to the runs in which subjects were instructed to use their left hand. Black contours show the target ROIs for each particular task. Colored arrows point at locations where ME-ICA resulted in greater consistency of activation across subjects inside (red) and outside (black) the target ROIs.

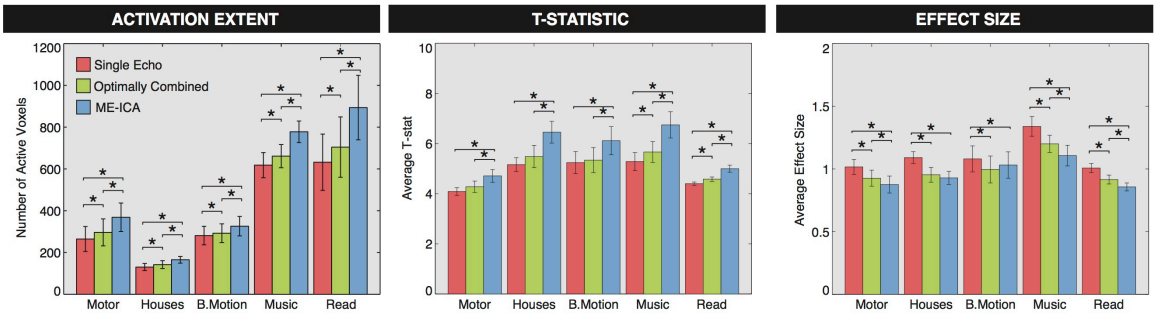


Figure 9. Performance analysis results for the rapid event-related dataset/per-task analyses. Post-hoc paired T-tests that reached significance (p<0.05) are marked with an asterisk.

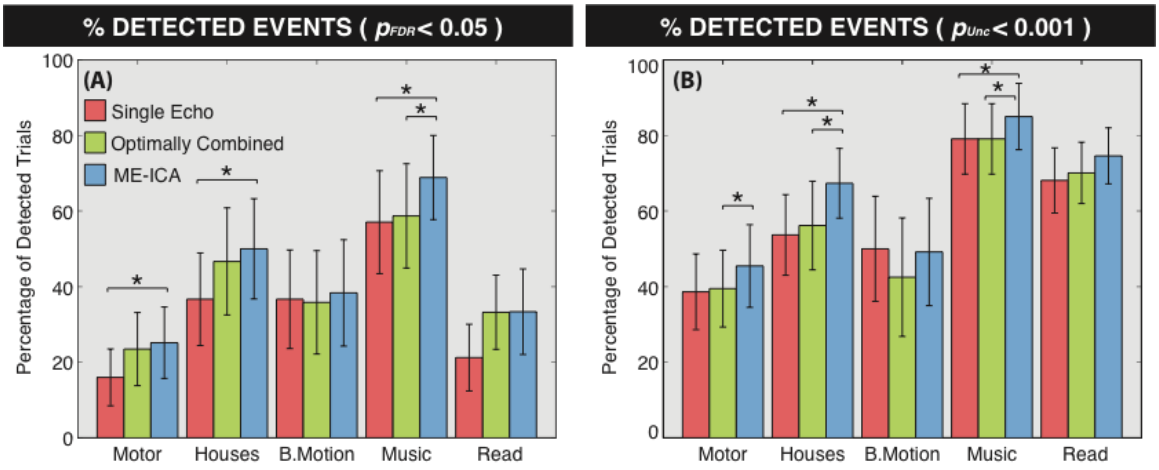


Figure 10. (A) Average percent of detected trials across tasks (p_{FDR} < 0.05) and pre-processing pipelines for the per-trial analysis of the rapid event-related dataset. Significant pair-wise T-tests are marked with an asterisk. Error bars represent standard error. (B) Same as (A) for p_{Unc} <0.001.

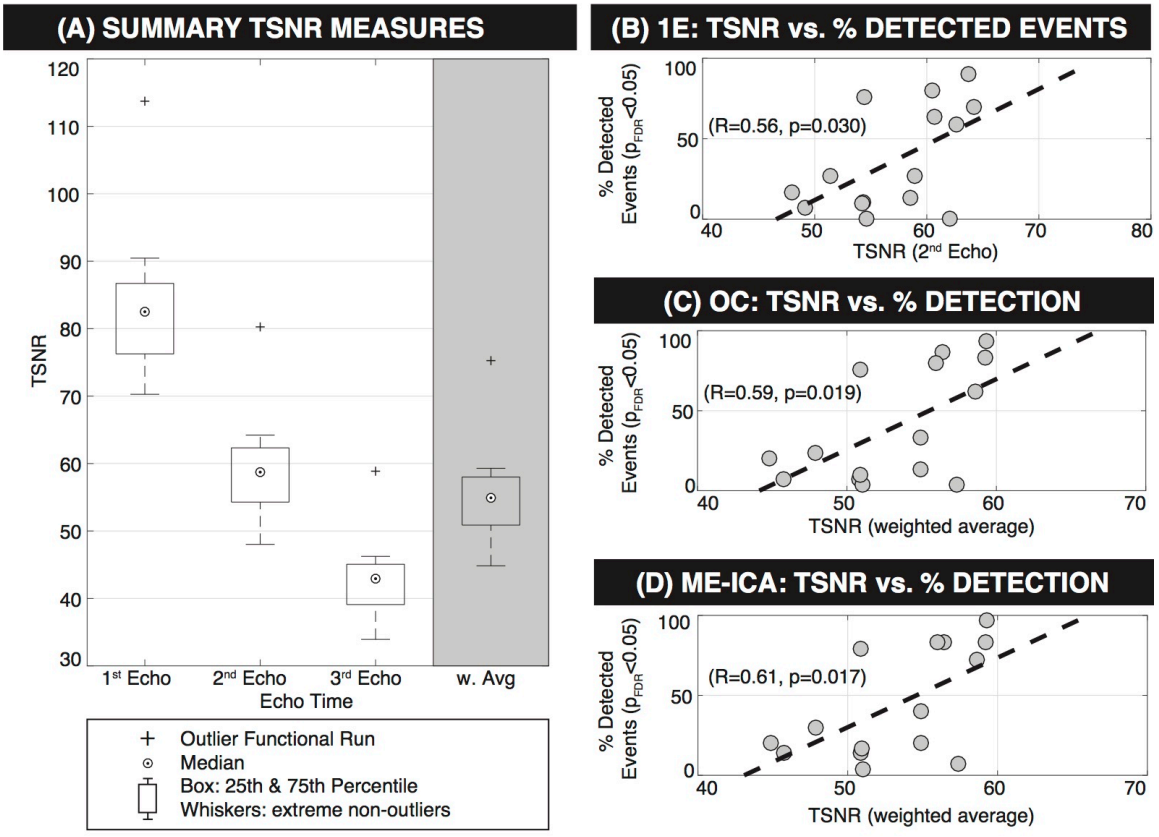


Figure 11. (A) TSNR per functional run and echo time. The weighted average TSNR across all three echoes is also presented (gray box). TSNR reported in terms of the

median (dot), 25–75% percentiles (box), and most extreme data points not considered outliers (dotted whiskers). In addition, outliers are marked with a (+) symbol. (B) Scatter plot of percent detected trials ($p_{FDR} < 0.05$) for the 1E pipeline versus temporal signal-to-noise ratio (TSNR). Each circle represents a different functional scan and the dashed line represents a least-squares linear fit to the data. (C) Same as (B) for the OC pipeline. (D) Same as (B) for the ME-ICA pipeline.

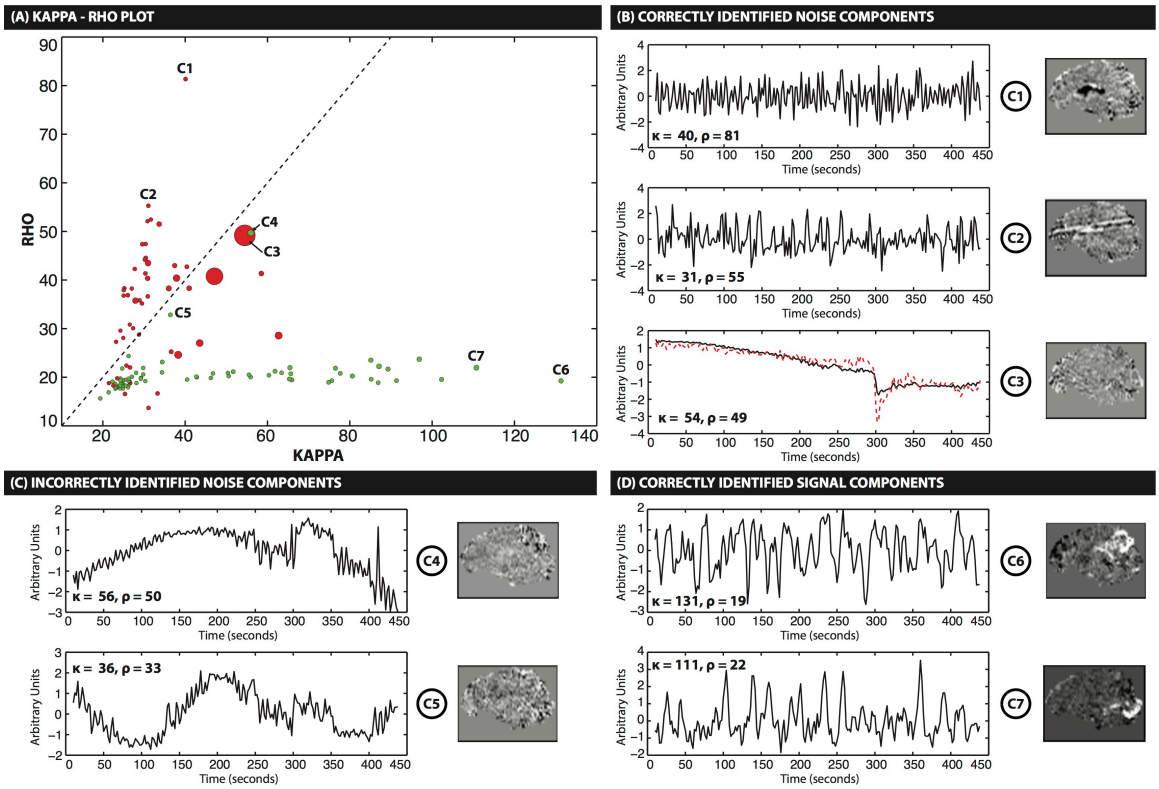
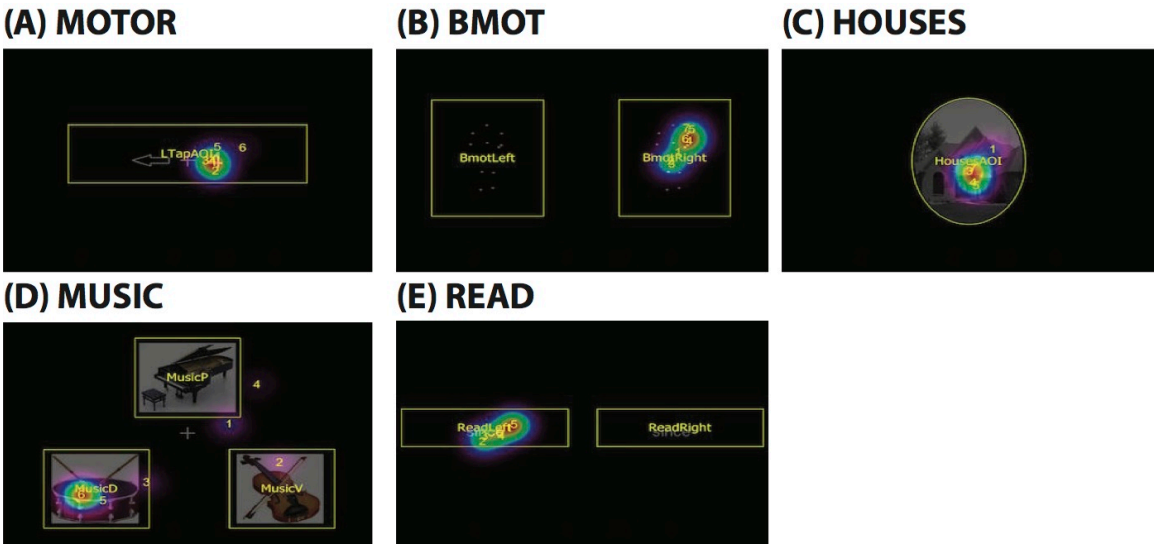
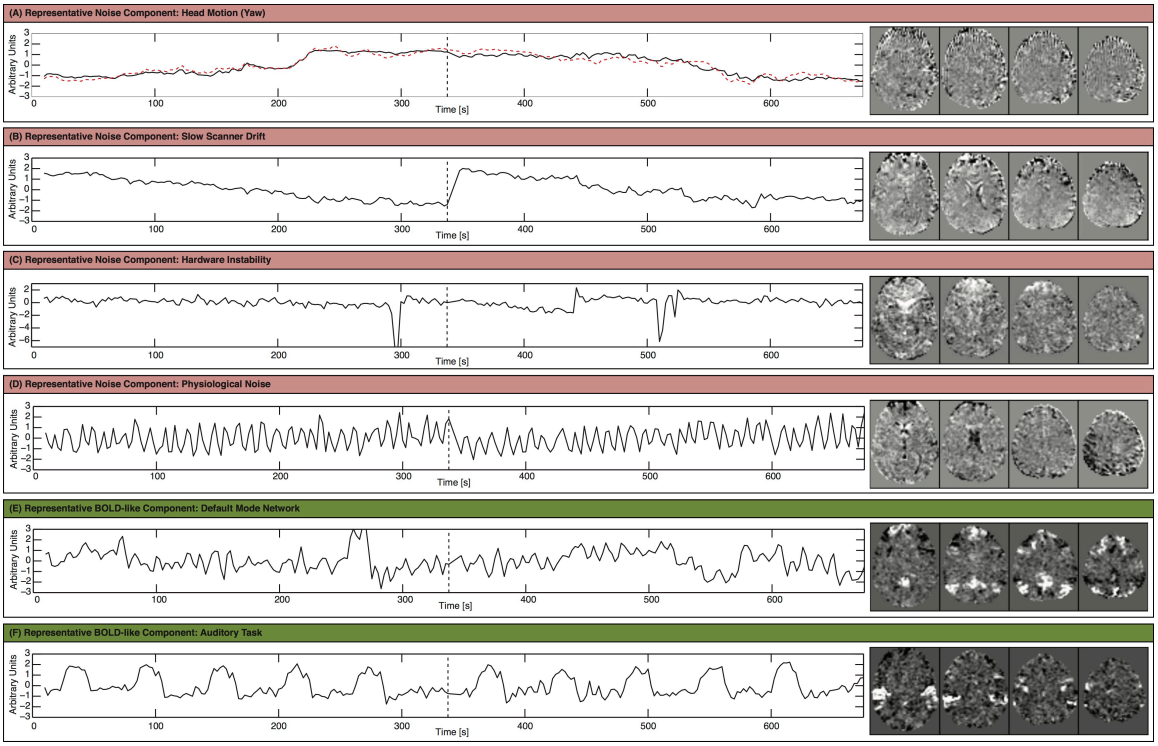


Figure 12. (A) Kappa – Rho spectrum for the ME-ICA decomposition of a representative subject that participated in the rapid event-related experiments. Each dot represents a component. The color of the dot indicates whether the component was marked as noise by the ME-ICA algorithm (red) or not-noise (green). The size of the dots is proportional to the amount of variance explained by the component. A dashed black line indicates locations in the plane where kappa equals rho. (B) Time series and spatial maps for three different noise component correctly identified as noise by ME-ICA. For the particular case of component C3, the component had high temporal correlation (r=0.87) with traces of head displacement in the AP direction (dashed red line). (C) Time series and spatial maps for two noise components incorrectly identified as BOLD by ME-ICA. (D) Time series and spatial maps for two BOLD components correctly identified by ME-ICA. For all components a label or black arrow indicates its location in the kappa-rho spectrum (A).

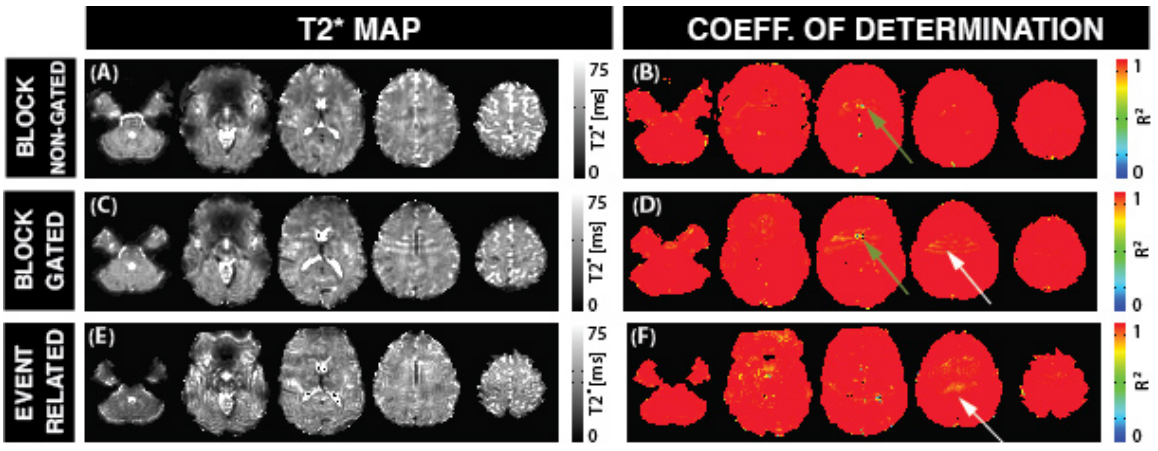
SUPPLEMENTARY FIGURE LEGENDS



Supplementary Figure 1. Visual field areas of interest (AOI) for the different tasks as defined in the OGAMA software for the analysis of eye tracking data. Areas of interest are delineated in yellow. Representative heat maps of visual focus for good subjects are superimposed. For tasks that show target stimuli in different hemifields during different trials (BMOT and READ) areas in both hemifields were defined, but only the correct one was used in each trial for the analysis. (A) AOI for the MOTOR task. (B) AOIs for the BMOT task. (C) AOI for the HOUSES task. (D) AOIs for the MUSIC task. In this case, the visual focus map indicates "drums" as the subject's response. (E) AOIs for the READ task.



Supplementary Figure 2. Representative ICA components detected by ME-ICA for the block design datasets. In all instances the component's time series are presented on the left and the corresponding spatial map on the right. All time series have been normalized to a mean of zero and a standard deviation of one. The transition between functional runs is marked with a vertical dashed black line. (A) Representative noise component associated with head motion artifacts. In addition to the component's time series we also present traces of rotational motion (dashed green line) that correlated heavily with this component. (B) Representative noise component associated with slow scanner drift. (C) Representative noise component associated with hardware instabilities. (D) Representative noise component associated with physiological noise. (E) Representative BOLD-like component depicting the default mode network. (F) Representative time series associated with the auditory task demanded from the subjects during these scans.



Supplementary Figure 3. Representative static T2* maps and associated goodness of fit maps as measured by the coefficient of determination (R^2). (A) Static T2* map for a representative dataset from the non-gated block design experiments. (B) Goodness of fit associated with the map in (A). (C) Static T2* map for a representative dataset from the gated block design experiments. (D) Goodness of fit associated with the map in (C). (E) Static T2* map for a representative dataset from the event-related experiments. (F) Goodness of fit associated with the map in (E). For all datasets we observe good fits, with the exception of some voxels in dorsal prefrontal regions affected by signal dropout at longer echo times (green arrow), and a few artifactual bands in more superior slices (white arrow).