A Paradigm Change in Functional Brain Mapping: Suppressing the Thermal Noise in fMRI

Vizioli, L.[*,1], Moeller, S.[*,1], Dowdle, L.[1], Akçakaya, M.[1,2], De Martino, F.[1,3], Yacoub, E.[1], Ugurbil, K[1].

[*] Equal first authorship

[1] Center for Magnetic Resonance Research (CMRR), University of Minnesota, Minneapolis, MN 55456

[2] Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, Minnesota

[3] Faculty of Psychology and Neuroscience, Department of Cognitive Neurosciences, Maastricht University, Maastricht, the Netherlands

Correspondence can be addressed to:
Kamil Ugurbil, ugurb001@umn.edu; Luca Vizioli, luca.vizioli1@gmail.com.

Main text wordcount (i.e. introduction, results, discussion and conclusion): 3487
Abstract wordcount: 151
Number of References: 43 for the main text; 49 Total

Abstract

Functional magnetic resonance imaging (fMRI) has become one of the most powerful tools for investigating the human brain. However, virtually all fMRI studies have relatively poor signal-to-noise ratio (SNR). We introduce a novel fMRI denoising technique, which removes noise that is indistinguishable from zero-mean Gaussian-distributed noise. Thermal noise, falling in this category, is a major noise source in fMRI, particularly, but not exclusively, at high spatial and/or temporal resolutions. Using 7-Tesla high-resolution data, we demonstrate remarkable improvements in temporal-SNR, the detection of stimulus-induced signal changes, and functional maps, while leaving stimulus-induced signal change amplitudes, image spatial resolution, and functional point-spread-function unaltered. We also provide supplementary data demonstrating that the method is equally applicable to supra-millimeter resolution 3- and 7-Tesla fMRI data, different cortical regions, stimulation/task paradigms, and acquisition strategies. The proposed denoising approach is expected to have a transformative impact on the scope and applications of fMRI to study the brain.

Introduction

Since its introduction in 1992, functional Magnetic Resonance Imaging (fMRI)[1-3] based on Blood Oxygenation Level Dependent (BOLD) contrast evolved to become an indispensable tool in the armamentarium of techniques employed for investigating human brain activity and functional connectivity. As such, it has been the central approach engaged in major initiatives targeting the human brain, such as the Human Connectome Project (HCP)[4], UK Biobank project[5], and the BRAIN Initiative[6].

In all techniques employed in imaging biological tissues, the need for improving the spatiotemporal resolution is self-evident and fMRI is no exception. To date, this challenge has been addressed primarily by increasing the magnetic field strength, leading to the development of ultrahigh magnetic field (UHF) of 7 Tesla (7T)[7]. UHF increases both the intrinsic signal-to-noise ratio (SNR) of the MR measurement as well as the magnitude and the spatial fidelity (relative to neuronal activity) of the BOLD based functional images[7,8]. These UHF advantages have enabled fMRI studies with submillimeter resolutions in the human brain, leading to the functional mapping of cortical columns and layers, and other fine-scale organizations[7,9,10]. Such studies provide unique opportunities for investigating the organizing principles of the human cortex at the mesoscopic scale, and bridging the gap between invasive electrophysiology and optical imaging studies and non-invasive human neuroimaging.

Despite these successes, however, the SNR and functional contrast-to-noise ratio (fCNR) of fMRI measurements in general remain quite low. This limitation is a major impediment to expanding the spatiotemporal scale of fMRI applications as well as the utility, interpretation, and ultimate impact of fMRI data. One strategy to alleviate this limitation is to pursue even higher magnetic fields[11-13]. Another simple and commonly used approach is the implementation of longer scan sessions, though this approach is not always optimal, since cognitive responses are modulated by attention and internal mental states, which likely change over long scan sessions. Moreover, long scans represent a limitation for most patient studies, hindering the translational applications to clinical problems.

In this paper, we tackle this SNR and fCNR limitation in fMRI using a denoising technique – namely, Noise Reduction with Distribution Corrected (NORDIC) PCA – which operates on repetitively acquired MRI data and *removes all components which cannot be distinguished from zero-mean Gaussian distributed noise*.

What is considered “noise” in an fMRI time series is a complex question. Thermal noise associated with the MR detection[14,15], arising either from the electronics and/or the sample, is a potential noise source in fMRI and would classify as a zero-mean Gaussian distributed noise. The use of parallel imaging to accelerate image acquisition, as is commonly done in contemporary MR imaging, introduces a spatially non-uniform amplification of this “thermal” noise by the g-factor[16]. The conditions under which this noise becomes dominant in an fMRI time series depends on the static magnetic field strength, the voxel volume, and image repetition time (TR) used in the experiment, becoming more prominent at higher resolutions (i.e. smaller voxel volumes), short TRs, and/or lower magnetic fields[17]. It is the *dominant* contribution at ~0.5 μL voxel volumes (e.g. ~0.8 mm isotropic dimensions) typically employed in high resolution 7T fMRI studies; it remains the dominant contribution at 7T up to ~10 μL voxel volumes, gradually plateauing beyond that[17]. However, even with 3 mm isotropic resolution (i.e. 27 μL voxel volume) and relatively long TR acquisitions, thermal noise of the MR measurement at 7T was estimated to still be a significant contributor to fMRI time series[18]. At lower magnetic fields like 3T, where this type of noise becomes more conspicuous, and where typical fMRI resolutions employed are $\lesssim$3 mm, it would be a significant contributor in virtually all fMRI studies[17]. It is this noise component that the method introduced in this work aims to suppress in the fMRI data and *not* the structured, non-white noise caused by respiration, cardiac pulsation and spontaneous neuronal activity (e.g.[19] and references therein).

Results of detailed analysis presented in this paper using high resolution 7 Tesla data demonstrate that major gains are achievable with NORDIC in gradient-echo (GE) BOLD fMRI without introducing image blurring. Supplementary data presented also document that the method achieves similar gains in more conventional, supra-millimeter spatial resolution fMRI studies carried out at both 3T and 7T, and with alternative task/stimulus and data acquisition strategies. Data acquired from longer acquisitions and averaged were used to evaluate the impact of NORDIC compared to standard reconstruction algorithms. Based on the gains demonstrated, the approach is expected to substantially widen the scope and applications of fMRI in general, and high spatial and/or temporal resolution fMRI in particular.

Results

NORDIC is a novel denoising algorithm that operates on repetitively acquired, complex-valued data to identify and remove signal components which cannot be distinguished from Gaussian noise[20]. Detailed description of the approach is given in the Methods section. Briefly, in NORDIC, the data matrix is constructed from complex-valued images, mapping the image reconstruction specific, spatially varying noise to spatially non-varying noise using the g-factor of the parallel imaging reconstruction method. Subsequently, the noise component of the data matrix has zero-mean, independent and identically distributed (i.i.d.) entries; for any local Cassorati matrix in the fMRI series, the threshold for the Gaussian distributed noise is then given unambiguously as the largest singular value of a matched Casorati matrix with i.i.d. noise. This choice ensures that all components that are indistinguishable from zero-mean Gaussian noise are removed.

The fMRI images were acquired with gradient echo Simultaneous Multi Slice (SMS)/Multiband (MB) Echo Planar imaging (EPI)[21,22]. They were processed by two different types of image reconstructions, a reconstruction produced by the MR scanner and referred throughout this work as “Standard”, and a reconstruction derived from the raw k-space files produced by the scanner and processed through the method of NORDIC PCA and is referred to as “NORDIC”.

The visual stimuli paradigm followed that employed by Shmuel et. al. with minor modifications[23]. The stimuli consisted of retinotopically organized target and surround stimuli (Figure 1A) presented in alternating stimulus-on epochs separated by stimulus-off control epochs. Each “run” consisted of a total of six stimulus-on epochs, three each for target and surround stimuli. We acquired 8 experimental runs in 2 subjects (see Methods); 2 of these were used to identify the corresponding retinotopic representation of the target condition in V1, computed by contrasting the activation elicited by the target to that of the surround condition ($p<0.01$ uncorrected). This functionally defined region of interest (ROI), referred to as “target ROI” from here on, was subsequently used for all ROI confined analyses. The functional runs used to estimate the ROI were excluded from subsequent analyses (see Methods). Additional subjects were employed for the 3T and 7T data presented in the accompanying Supplementary Material.

NORDIC vs. Standard MR Images: Figure 1B shows an example slice for Standard and NORDIC reconstructed GE-SMS/MB-EPI images for both subjects, before any preprocessing for fMRI analysis was applied. The anatomical definition of the gray matter, gray/white matter

boundary, and venous blood vessels located within sulci (appearing as dark spots) emerge sharper and clearer in NORDIC images, especially in the central regions where g-factor noise

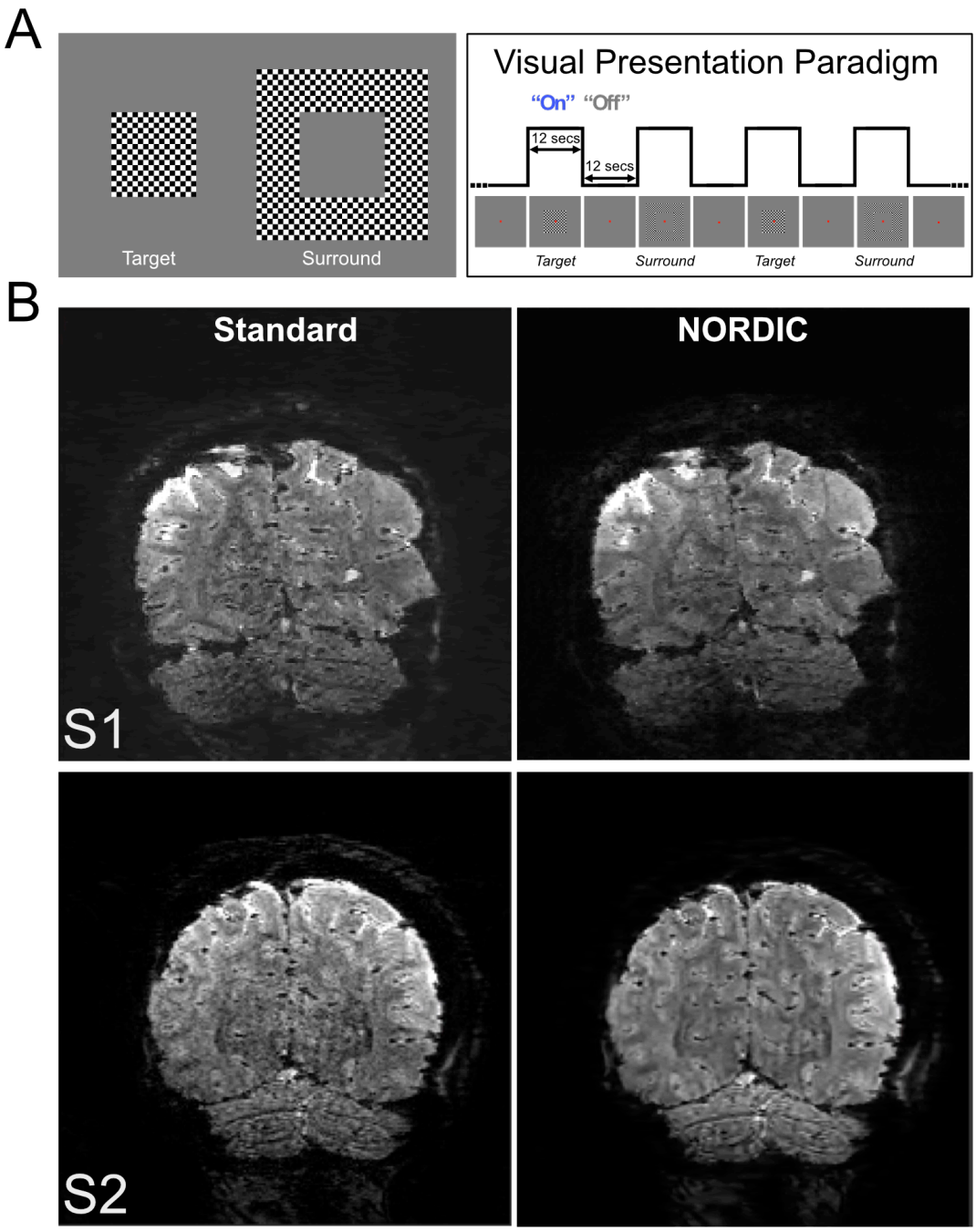


Figure 1. Panel A depicts the visual stimuli (left) used and a schematic of the visual presentation paradigm (right). Panel B shows an example slice from a single volume extracted from an fMRI time series for Standard (left column) and NORDIC (right column) reconstructions before any preprocessing, for two subjects S1 and S2.

amplification is particularly elevated due to the highly accelerated image acquisition employed. It is expected that NORDIC suppresses this noise, thus improving image quality, as observed (Figure 1B).

Figure 2A shows temporal SNR (tSNR) maps averaged across all eight runs for two exemplar slices (see also Supplementary Figure 1). The average tSNR across *all* the voxels in

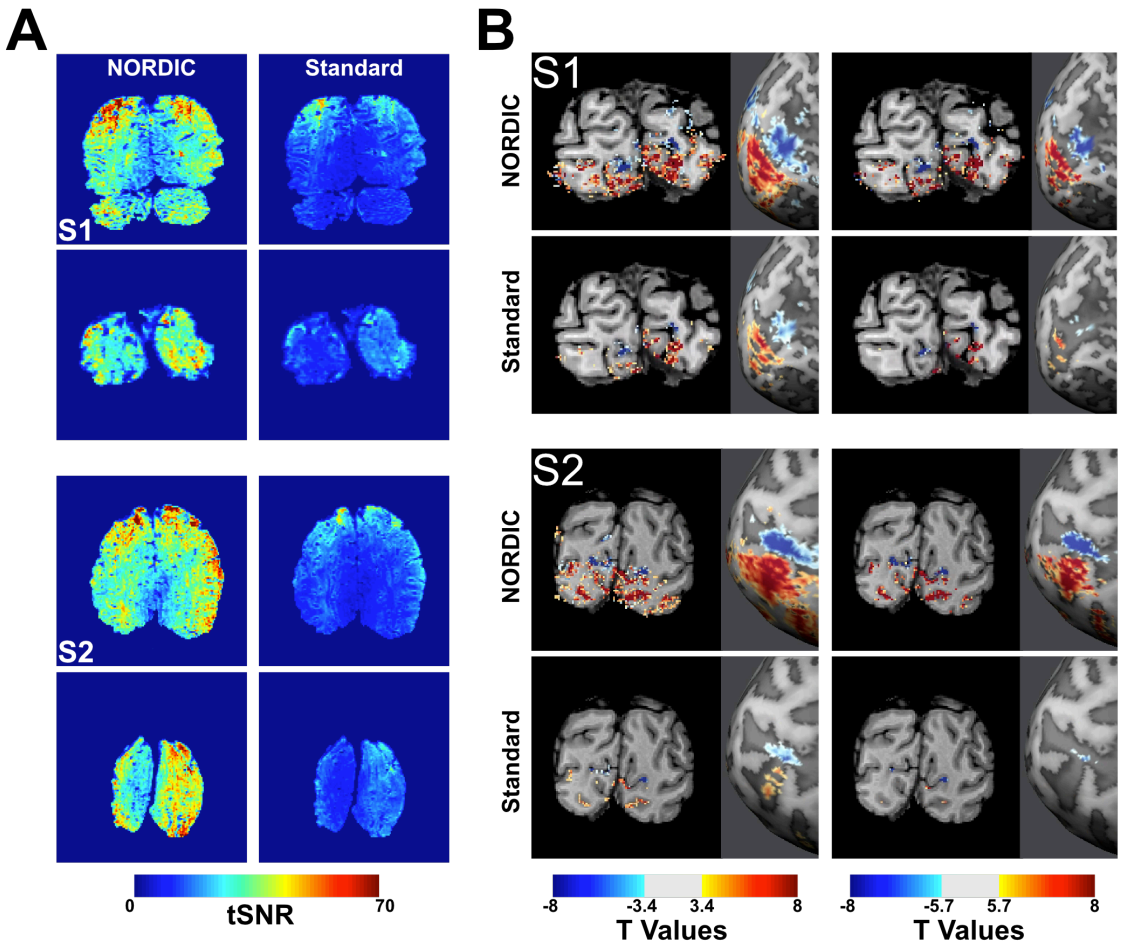


Figure 2. Panel A) Average (across all 8 runs) brain temporal SNR (tSNR) maps of 2 exemplar slices in both subjects (S1 and S2) for Standard (left) and NORDIC (right) reconstructions. The slices choses represent one of the anterior most slices in the covered volume, and an occipital slice that includes a portion of the ROI V1. Panel B) Examples of NORDIC and Standard reconstructed t-maps superimposed onto T1 weighted anatomical images for both subjects (S1 and S2) obtained from a single run. For each subject, the images on the top row represent NORDIC reconstruction and the lower row the Standard reconstruction. The t-maps were computed by contrasting the activation elicited by the target (red) *versus* that elicited by the surround (blue) condition for a single representative run in volume space (left) and on inflated brains (right). The two different columns show different thresholds, specifically, $t \geq |3.4|$ (left column in panel B) and $t \geq |5.7|$ (right column in panel B); these t-values correspond to $p<0.01$ (uncorrected) and $p < 0.05$ (Bonferroni corrected) for the Standard reconstruction.

the brain was more than 2-fold larger for NORDIC ($S1_{tSNR}$: 27.34±2.26 (std); $S2_{tSNR}$: 33.01±2.26 (std)) compared to Standard ($S1_{tSNR}$: 13.28±0.19 (std); $S2_{tSNR}$: 13.97±0.05 (std)) images. Paired sample t-tests carried out across all 8 runs, independently per subject, indicated that for both subjects, the average tSNR for NORDIC was significantly larger ($p<0.01e^{-5}$) than that estimated for Standard images.

Functional Images: Figure 2B depicts t-maps of a single representative run computed by contrasting the target *versus* the surround condition for the two reconstructions at two different thresholds. The t-maps are displayed for a single slice as well as on the inflated surface of one hemisphere. The regions responding to the target and the surround stimuli appear in red and blue, respectively. Given that GLM t-values can be conceptualized as representing approximately the stimulus induced signal change amplitude divided by the noise in the fMRI time series, a larger extent of activation is expected and observed in the t-thresholded maps of NORDIC compared to Standard (Figure 2B).

To quantify the difference evident in Figure 2B across reconstructions, we compared a single run processed with NORDIC against the concatenation of multiple runs obtained with the Standard reconstruction (see Methods). Figure 3A illustrates these functional maps on the inflated surface of one hemisphere, thresholded at $t \geq |5.7|$ for the two subjects. At the same t-threshold, the extent of activation achievable with a single NORDIC run was comparable to that obtained by concatenating 3 to 5 Standard runs (Figure 3A). Similar results are presented in the supplementary section for supra-millimeter 3T and 7T fMRI data obtained with visual stimulation and face recognition paradigms (Supplementary Figures 5-7) and for 0.8 mm 7T data obtained with auditory stimulation (Supplementary Figure 9).

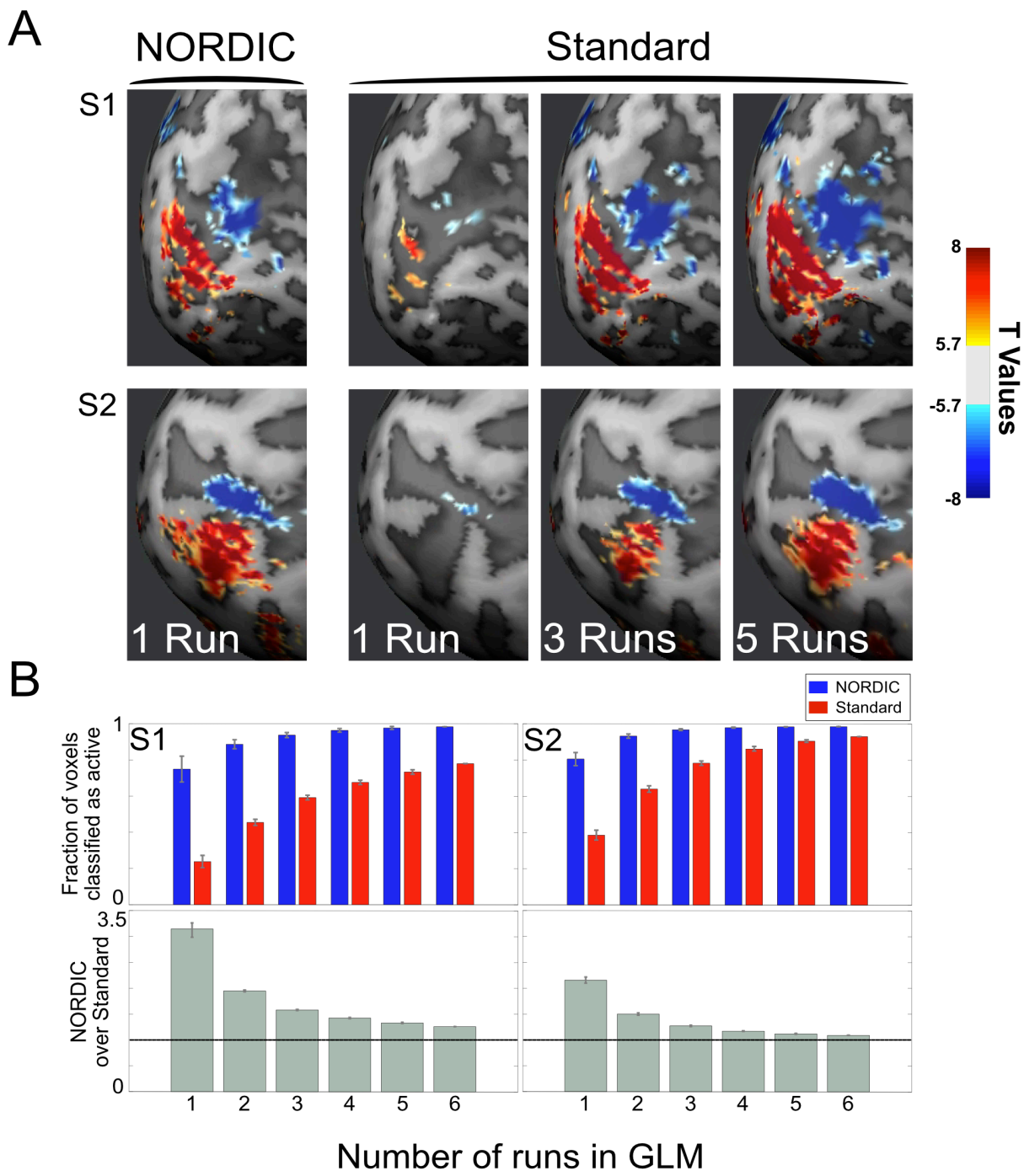


Figure 3. Panel A shows functional images as t-maps (target > surround) thresholded at t ≥ |5.7| for a single NORDIC processed run (left most panel) and for 1, 3 and 5 Standard processed runs combined, for subject 1 (S1, top row) and subject 2 (S2, lower row). Panel B shows the *fraction* of the voxels with t ≥ 5.7 within a predefined ROI that defines the left retinotopic representation of the target stimulus; the results (panel B) were calculated as the bootstrap mean across all run combinations for GLMs computed on 1, 2, 3, 4, 5 and 6 runs (see Methods). Blue and Red bars represent NORDIC and Standard reconstructions, respectively. Panel B lower row is the ratio of the two.

Within the target ROI, the fraction of voxels with $t \geq 5.7$ were compared for 1 to 6 runs processed with NORDIC and Standard reconstructions (Figure 3B); this was calculated as the bootstrap mean across all run combinations (see Methods) of GLMs computed on 1 to 6 runs (arranged on the x-axis). With NORDIC, approximately 3 to 4 runs were needed to achieve $t \geq 5.7$ for approximately all voxels within the target ROI; in contrast, with the Standard reconstruction, even 6 runs failed to meet or exceed this t-threshold for the entire target ROI, especially in subject 1. In addition, data presented in Figure 4 obtained with linear mixed models carried out independently per subject demonstrate that within the target ROI, the t-values were significantly larger for NORDIC ($p<0.05$) than its Standard counterpart (Figure 4B and 4C) for both subjects.

Figure 4D presents data on the percent signal change (PSC) within the target ROI as the mean of all voxels (Figure 4D) and at the single-voxel level (Figure 4E) within the target ROI. The stimulus induced PSC was highly comparable across reconstruction types (Figures 4D and 4E; see also Supplementary Figure 2); linear mixed models carried out independently per subject (with the individual runs as random effect) showed no significant ($p>0.05$, Bonferroni corrected) differences in PSC amplitudes across reconstructions for all runs.

Figure 4F depicts the standard deviation (20% trimmed mean across voxels within target ROI[24]) computed amongst PSC betas elicited by a single presentation of the target stimulus within a run. As shown by both paired sample t-test ($p<0.05$) and 95% bootstrap confidence interval (carried out by sampling with replacement of the individual runs), this metric was found to be significantly larger for Standard than NORDIC for all subjects, indicating greater stability of NORDIC PSC single trial estimates among the different stimulus epochs within a run. These observations were further confirmed in Figure 5.

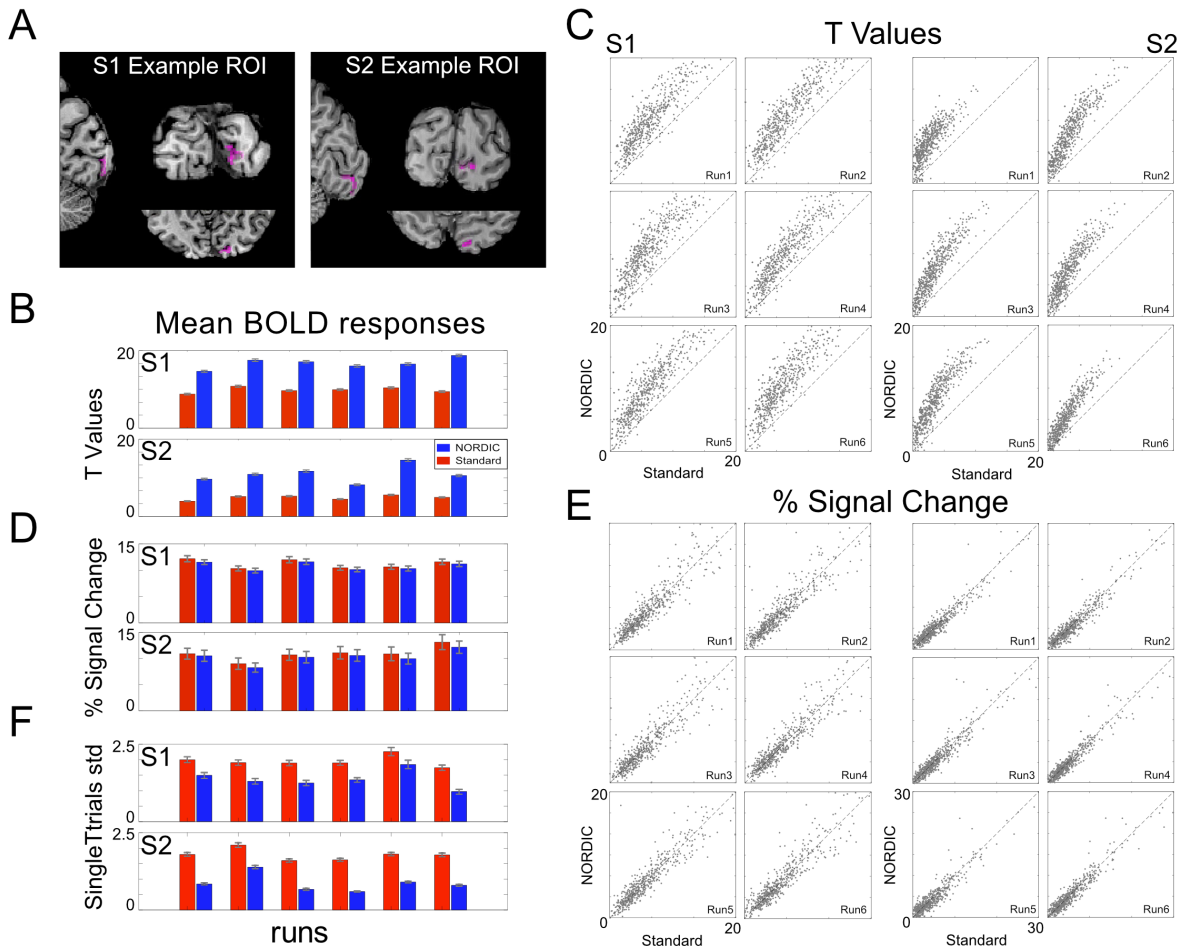


Figure 4. Panel A shows the target ROI, representing the left retinotopic representation of the target in V1 for both subjects in all 3 planes. Panel B shows the single-run average beta weights (transformed into percent signal change) across all voxels within the target ROI elicited by the target stimulus. In blue are the responses for NORDIC images, while in red are those for Standard images for all runs (arranged on the x axis). Error bars represent standard error across all voxels within the ROI. Panel C: Single run, single voxel scatterplots for the beta percent signal change responses to the target condition, for Standard (x axis) and NORDIC (y axis). Panel D: same as panel B, but for average t-values (activity elicited by the target > 0) within the predefined ROI. Panel E: same as panel C for t-values (activity elicited by the target > 0). Panel F shows the single-run standard deviation (averaged across all voxels within the ROI) computed across single trial PSC beta estimates elicited by the target condition.

The equivalency of PSC amplitudes for NORDIC and Standard reconstructions are illustrated using images in Figure 5A (top two rows). Figure 5 also presents the results of an hold-out data analysis carried out with PSC estimates; for this, we estimated GLM model parameters in one run and assessed the precision with which these parameters predicted the PSC in all other runs at a single voxel level. The precision of PSC estimates, computed as cross-validated R2 for single run GLMs was higher (Figure 5A, third row) for the NORDIC compared to the Standard reconstruction. T-test carried out across cross-validations folds showed that within the target ROI

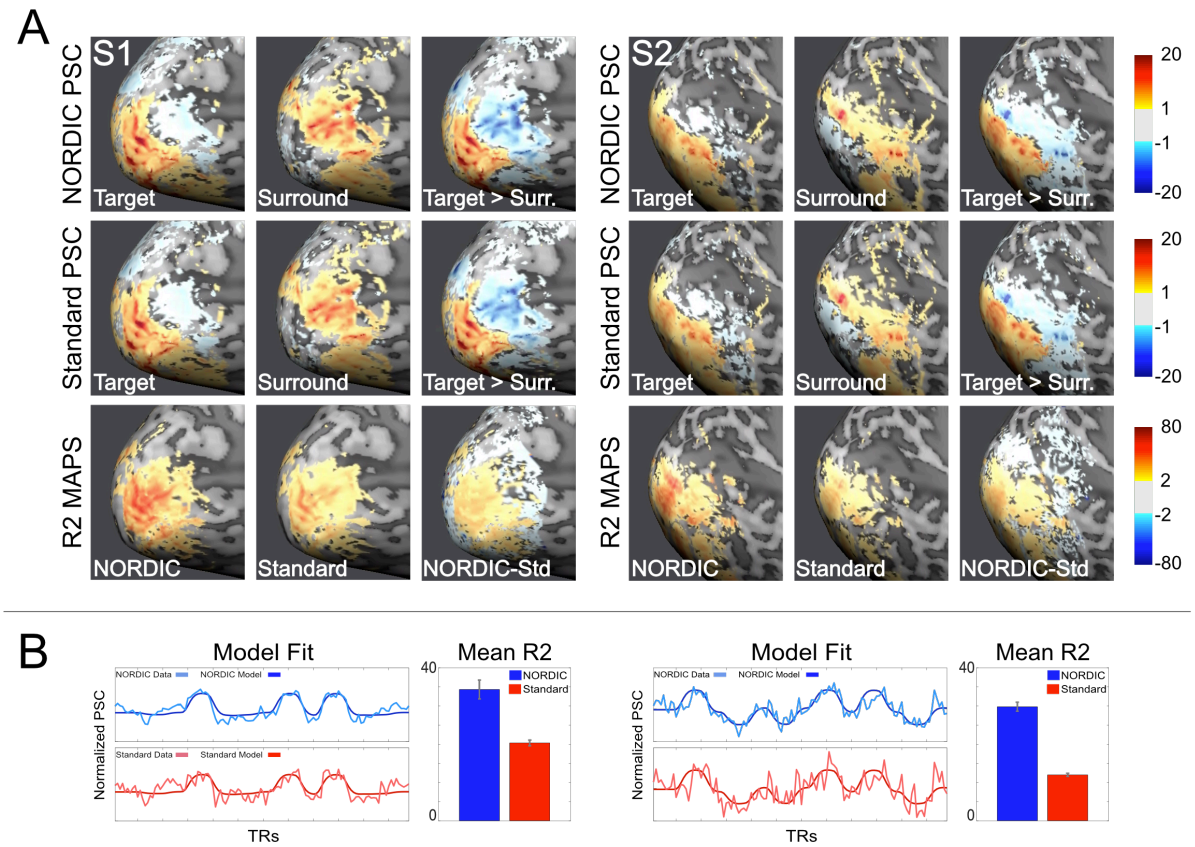


Figure 5. Panel A: Average percent signal change (PSC) maps and cross-validated R2 for both subjects S1 and S2. The top 2 rows in Panel A show the average (across runs) PSC maps elicited by the target (left), surround (middle) and their contrast target>surround (right), for NORDIC and Standard reconstructions, respectively. As evident by these images, PSC amplitude and extent of ethe extend of stimulus induced signal change is comparable across reconstructions (see also Supplemental Figure S2). The 3rd row in panel A shows the average (across folds) cross-validated R2 maps for NORDIC (left), Standard (middle) and their difference (right). *Only in the relevant portion of the cortex for the stimulus used* (i.e. areas where stimulus induced BOLD activity is expected, as indicated by the PSC maps) the R2 maps show higher precision of PSC estimates for NORDIC images. Panel B: For each subject, left column show a single run of NORDIC and Standard BOLD time-courses for a target-selective voxel (lighter lines) and its prediction estimated on a separate run (darker lines). These plots highlight a closer correspondence between model prediction and empirical time-courses for NORDIC data, as summarized by the significantly larger ($p<.01$) target ROI average cross-validate R2 (bar plots; error bars indicate standard errors across cross-validation folds).

average R2 (see Methods) was significantly (p<0.01) higher for NORDIC (S1: NORDIC mean R2=34.34 (ste=2.24); Standard mean R2=20.36 (ste=0.95); S2: NORDIC mean R2=29.94 (ste=0.97); Standard: mean R2=12.04 (ste=0.56) and Figure 5B, bar graphs), indicating again higher precision of PSC estimates for the former and their stronger predictive value.

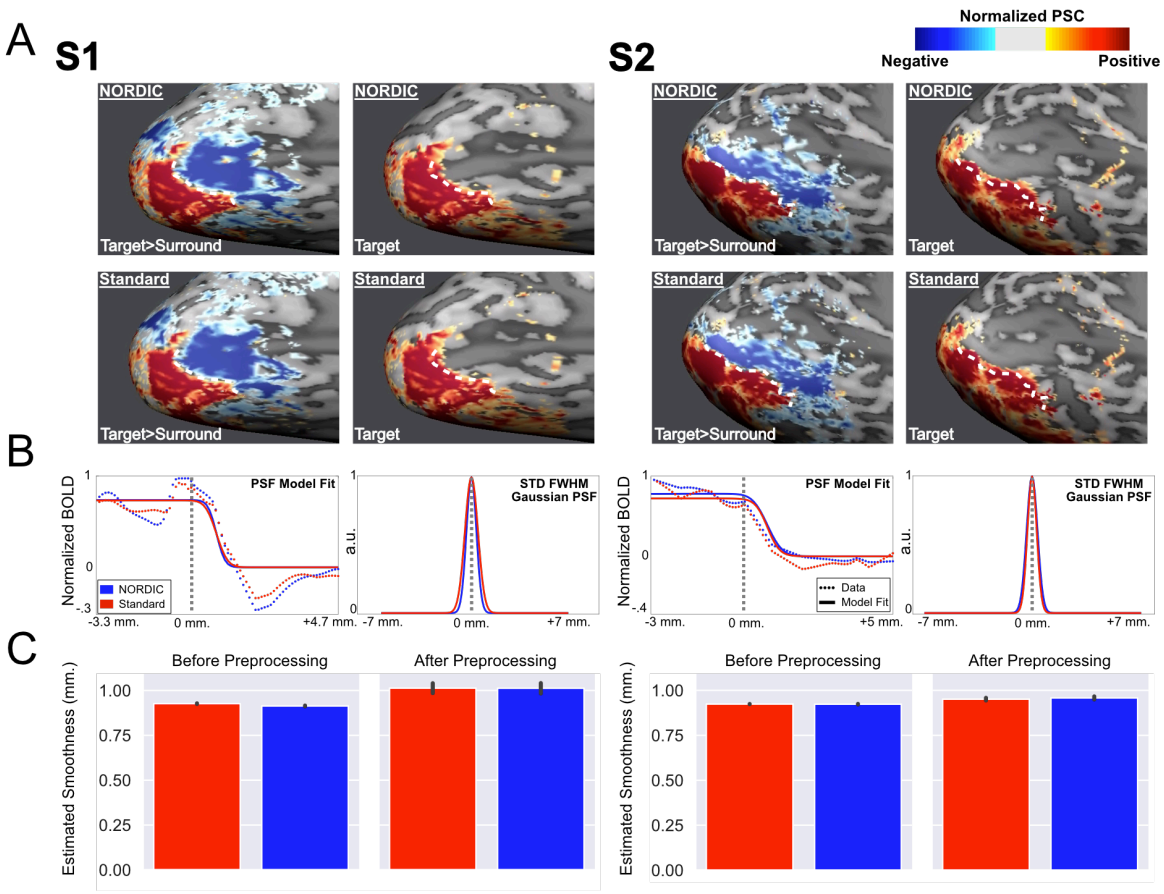


Figure 6. Functional point spread function (PSF) (top 3 rows) and global image smoothness (bottom row). for subjects S1 and S2. Top row shows NORDIC normalized beta PSC maps for *differential* mapping target (in red) > surround (in blue), and the target only (left) *single-condition* image. The black dotted line is determined in the *differential* image as the "boundary" between the two stimulations. The same black dotted line is also superimposed on the target only PSC map where PSC values greater than zero but decreasing in magnitude progressively away from this "boundary" posteriorly are seen. The functional PSF is calculated from this spread in PSC beyond the "boundary". The second row is identical to the first row, but obtained from Standard reconstruction data. 3rd row, left panel for each subject: the PSC magnitude changes (normalized to the highest value) along traces perpendicular to the "boundary" intensity are displayed on the left panel as the average (across traces and runs). The model fits (solid line) and data (dotted line) are shown for both the NORDIC (blue) and Standard (red) reconstructions. The vertical grey dotted line represents the "boundary" as derived by the differential maps. 3rd row, right panel portrays the full width half max (FWHM) standard deviation of the gaussian kernel that was convolved with a step function to model functional PSF (see Methods). Bottom row: Global smoothness of images used for the fMRI time series for Standard (red) and NORDIC (blue), before (left panel) and after preprocessing related interpolations.

Figure 6 shows the functional point spread (PSF) measurements on the cortical surface calculated following previous work[23] using NORDIC and Standard images (see Methods): briefly, the approach defines the boundary between the target and the surround stimuli as those voxels showing a *differential* functional response close to 0 (Figure 6A, left column for each subject). Along traces drawn orthogonal to this boundary, the functional response amplitudes are then measured in the *single condition* maps and subsequently quantified by fitting a model consisting of a step-function (representing infinitely sharp PSF) convolved with a Gaussian[23] (Figure 6B). The full width half max (FWHM) of the Gaussian represents the functional PSF[23]. With NORDIC, the average PSFs (across traces) were 1.04 mm (std: 0.19) and 1.22 mm (std: 0.51), for subject 1 and 2, respectively; the average PSFs for the Standard were 1.14 mm (std: 0.16) and 1.15 mm (std: 0.11). Paired sample t-tests carried across the 8 runs showed no significant differences ($p>0.05$) in functional PSF amongst reconstruction types for either subject.

In addition, we estimated the global smoothness of individual GE-SMS/MB-EPI images in the fMRI time series using AFNI (3dFWHMx function)[25], with automatic intensity-based masking derived from the median image of each run. The spatial autocorrelation was estimated using a Gaussian + monoexponential decay mixed model to account for possible long-tail autocorrelations found in fMRI data. The FWHM from this mixed model estimate, averaged over subjects, before and after data preprocessing (see Methods) for the Standard reconstruction was 0.925±0.002mm, and 0.982±0.039mm, respectively; for the NORDIC reconstruction these values were 0.918±0.005 and 0.985±0.036mm (Figure 6C). Paired samples tests find that NORDIC reconstructed images were significantly ($p<0.05$ Bonferroni corrected) less smooth than standard before processing. After processing, no significant difference was observed in smoothness ($p>0.05$ Bonferroni corrected), demonstrating that NORDIC did not introduce any image blurring.

Discussion

We have previously described NORDIC and its application in diffusion weighted imaging (dMRI) employed for tractography[20]. Like fMRI, dMRI has an inherent assumption of redundancy explicit in repetitive acquisitions[26]. The possibility of denoising dMRI data using PCA decomposition was pointed out as early as 2013[27], and gained increasing attention with the recently introduced Marchenko-Pastur Principle Component Analysis (MPPCA) method[28]. It is difficult, however, to identify precisely the components that have been removed for a finite series with unknown low-rank in the MPPCA approach. These and other challenges were tackled by

NORDIC, which yields a parameter-free objective threshold for identifying i.i.d, zero-mean Gaussian contributions based on random matrix theory. In dMRI applications, NORDIC was shown to yield superior results to MPPCA[20]. In this paper, we demonstrate that it provides major gains for fMRI.

For task- and/or stimulus-based fMRI, temporal signal fluctuations that are unrelated to the task or stimulus presentation represent "noise" in an fMRI time series, and fCNR is simply the product of tSNR and the stimulus/task induced percent signal change (e.g. Appendix 1 in[7]). Certainly, thermal noise of the MR detection[14,15,29], including the spatially non-uniform g-factor noise amplification[16] associated with accelerated acquisitions, negatively impacts tSNR. Temporal fluctuations induced by physiological processes of respiration and cardiac pulsation[30] represent another source of tSNR degradation; these "physiological" fluctuations are non-white in nature[31] and are detectable in the fourier transform of the fMRI time series when the sampling rate is sufficiently fast (e.g.[32]). Finally, spontaneous, spatially correlated fluctuations in the signal, ascribed to functional networks in resting state fMRI[33] also represent a non-gaussian "noise" that impacts fCNR[18,34,35] for stimulus/task fMRI.

Other than the thermal noise of MR detection, all aforementioned sources of signal fluctuations in an fMRI time series are proportional to signal magnitude[36-40]; as such, they become dominant when a voxel's signal (which is proportional to voxel volume) is large compared to thermal noise, as encountered at low spatial resolutions (i.e. large voxel volumes), and especially at high magnetic fields[17,40]. Numerous methods have previously been introduced to remove this particular confound from fMRI data (e.g.[19,31] and references therein). However, thermal noise is a discernable and significant contribution even at resolutions conventionally used in contemporary 3 or 7 Tesla studies (e.g. 1 to 2 mm isotropic or equivalent) and becomes increasingly more dominant at higher resolutions. It is this thermal component of fMRI noise that the present method selectively targets for removal. As a result, tSNR is significantly better throughout the brain for NORDIC compared to Standard, non-denoised data (Figures 2A and Supplementary Figure 1), which translates into substantial increase in the magnitude of the t-values (Figure 4, and Supplementary Figures 5 and 8).

FFT power spectrum of the fMRI times series from NORDIC and Standard processed data are shown in Supplementary Figure 4; as expected, only a broadband decrease in the magnitude of the spectrum is seen, not impacting the various peaks detected at specific frequencies in the spectrum. In fact, respiration related peak at ~0.27 Hz in the spectrum becomes even more visible in NORDIC data.

T-values are a useful metric in evaluating the functional activation studies. Denoising algorithms inherently alter the dimensionality of the data and, consequently, the degrees of freedom (DFs) of GLM computations. GLM's DFs are crucial in computing p-values, though the correct computation of DFs for an fMRI time series is debated[41]. However, as shown in Eq. 2 (see Methods), DFs do not impact the related t-values. As such, we chose to threshold our activation maps based on t- rather than p-values, although we give the equivalent p value as a reference for the Standard reconstruction.

At the same t-value threshold, the extent of activation, i.e. voxels showing stimulus invoked signal changes that pass the t-threshold, is considerably larger for NORDIC for a single run (Figures 2 and 3; Supplementary Figures 5-7 and 9). For a given t-threshold, the activation maps produced by concatenating 3 to 5 runs of the Standard data approach become extremely similar to that obtained by NORDIC in a single run (Figure 3). This was also observed for supra-millimeter 3T and 7T data obtained in visual and fusiform face areas respectively, and for high resolution 7T fMRI data obtained in the auditory cortex (Supplemental Figures 5-7, 9),

Moreover, while NORDIC denoising does not impact the amplitude of evoked PSC responses (Figures 5, 4D, 4E and Supplementary Figure 2), it does improve the trial-to-trial precision of its single-voxel estimates. The stability of single-trial stimulus evoked response within a given run was over ~2 times better (i.e. lower in standard deviation) for NORDIC reconstructed data than its Standard counterpart (Figure 4F) for all runs and subjects ($p<0.05$ Bonferroni corrected). Cross validates R2 computed on NORDIC data were generally larger in relevant portions of the cortex indicating a selective improvement in the trial-to-trial precision of the PSC amplitude estimates (Figure 5; Supplementary Figure 3), consistent with the observation that NORDIC better estimates the stimulus evoked responses in shorter runs in fMRI studies. Single trial responses represent a challenging SNR starved scenario and capturing them accurately with low single-trial variance is a highly desirable, yet seldomly achievable feat, especially in submillimeter resolution fMRI.

One of the most important features of the NORDIC is its ability to preserve spatial precision of the individual images of the fMRI time series, as well as the functional response. Thermal noise of the MR detection process can and often is suppressed with spatial filtering, which smooths (i.e. blurs) the images, increasing the SNR and consequently the tSNR when thermal noise contribution to tSNR is significant[42]; this improves the t-statistics (Supplementary Figures 5 and 8) in an fMRI experiment. However, the resultant spatial blurring leads to a loss of spatial precision, which is highly undesirable. Nevertheless, spatial-smoothing is almost always

employed in fMRI due to the inherently low sensitivity of the fMRI technique even though it degrades the effective spatial resolution. The NORDIC approach, on the other hand, suppresses thermal noise and has the same impact on t-statistics as spatial-smoothing (Supplementary Figures 5 and 8) but *without* spatial blurring of either the individual images themselves (Figure 6C; also see Supplemental Figure 10 and related supplemental discussion) or the functional PSF estimates in the visual cortex (Figure 6A and 6B), yielding PSF values that are consistent with previous reports[23,43].

Conclusion

In this paper we demonstrate an fMRI denoising approach to remove thermal noise inherent in the MR detection process, and markedly improve some of the most fundamental metrics of functional activation detection while crucially preserving spatial and functional precision. We demonstrate its efficacy for 7T mapping at high spatial resolution, typical of the rapidly increasing number of contemporary studies targeting mesoscopic scale cortical organizations in the human brain as well as for 3T and 7T fMRI studies using the more commonly employed supra-millimeter spatial resolutions targeting different cortical regions activated by different stimuli and tasks. Importantly, as it specifically acts on Gaussian distributed noise, NORDIC is *complementary* as well as *beneficial* to denoising algorithms that primarily focus on structured, non-white noise removal. The cumulative gains are expected to bring in transformative improvements in fMRI, permitting more precise quantification of functional responses, robust studies of mesoscopic scale organizations even at resolutions that have been unreachable to date (e.g. 0.5 mm isotropic (see Supplemental Figure 10)), faster acquisitions rates, and/or significantly shorter scan times.

Materials and Methods

NORDIC PCA

NORDIC PCA is a denoising algorithm that utilizes the locally low rank (LLR) properties of an image series, operating on complex-valued data to remove noise that can be characterized as zero-mean i.i.d. Gaussian. Principles of NORDIC denoising are described below and adapted from [20].

Locally low rank Model: We let $\boldsymbol{m}(\mathbf{r}, t) \in \mathbb{C}^{\mathrm{I}_1 \times \mathrm{I}_2 \times \mathrm{I}_3 \times \mathbf{N}}$ denote a complex-valued volumetric fMRI image series following an accelerated parallel imaging acquisition, where **r** indicates the 3D spatial location, an *N* is the number of temporal samples. For LLR modeling, a $k_1 \times k_2 \times k_3$ patch whose top left corner is located at a given voxel $\mathbf{r_0}$ is selected. To reduce the complexity of notation, the dependence on $\mathbf{r_0}$ will be assumed implicit, and arbitrary patches will be considered. For a given time point $\mathrm{t} \in \{1, \cdots, N\}$, the patch will be vectorized as $\mathbf{y}_{\boldsymbol{t}}$, and a Casorati matrix $\mathbf{Y} = [\mathbf{y_1}, \cdots, \mathbf{y}_{\boldsymbol{t}}, \cdots, \mathbf{y_N}] \in \mathbb{C}^{\mathrm{M} \times \mathbf{N}}$ will be generated with $M = k_1 \times k_2 \times k_3$. The goal of denoising is to estimate the underlying data Casorati matrix $\mathbf{X}$, based on the noisy observed matrix $\mathbf{Y}$, with

$$\mathbf{Y} = \mathbf{X} + \mathbf{N} \qquad \text{(eq.1)}$$

where $\mathbf{N} \in \mathbb{C}^{\mathrm{M} \times \mathbf{N}}$ is additive Gaussian noise. LLR models assume that each patch across the fMRI series has an underlying data Casorati matrix $\mathbf{X}$ with a low-rank representation. LLR methods typically implement this representation by singular value thresholding (SVT). In SVT, singular value decomposition is performed on $\mathbf{Y}$ as $\mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^H$, where the entries of the diagonal matrix $\mathbf{S}$ are the ordered singular values, $\lambda(j)$, $j \in \{1, \cdots, N\}$. Then the singular values below a threshold $\lambda(j) < \lambda_{thr}$ to $\lambda(j)$=0 while the other singular values are unaffected. Using this new diagonal matrix $\mathbf{S}_{\lambda_{thr}}$, the low-rank estimate of $\mathbf{Y}$ is given as $\boldsymbol{Y_L} = \mathbf{U} \cdot \mathbf{S}_{\lambda_{thr}} \cdot \mathbf{V}^H$. The patches arising from these thresholded Casorati matrices are then combined by averaging overlapping patches or by using non-overlapping patches to generate the denoised image series $\boldsymbol{m}^{LLR}(\mathbf{r}, \tau)$.

Noise Model and Threshold Selection: The distribution of the singular values of a random noise matrix $\mathbf{N}$ is well-understood if its entries are i.i.d. zero-mean, and is asymptotically characterized by the Marchenko-Pastur distribution. In NORDIC, the data matrix $\mathbf{Y}$ is constructed to ensure that the noise matrix component of $\mathbf{Y}$ has this property, as follows: 1) Images are processed as complex-valued maintaining a zero-mean Gaussian noise following image

reconstruction, 2) Spatially varying noise in the reconstructed images are normalized by the g-factor maps of the parallel imaging method to lead to spatially identical noise, and 3) The threshold is selected explicitly based on the noise spectrum distribution.

To this end, SENSE-1 combination[44] using coil sensitivity profiles generated with ESPIRIT algorithm is applied to the slice-GRAPPA reconstructed simultaneous multislice (SMS)/Multiband (MB) accelerated dataset[45] to maintain Gaussian noise[46]. Then g-factor maps, $g(\mathbf{r})$ are calculated building on the approach outlined in [47] for GRAPPA reconstructions[45] with the same sensitivity profiles used for image reconstruction. Signal scaling is performed as s $\boldsymbol{m}(\mathbf{r},\mathrm{t})/g(\mathbf{r})$ to ensure zero-mean and spatially identical noise in a given patch. Furthermore, the independence requirement of the entries of noise component matrix is satisfied by choosing a sufficiently small patch size so that no two voxels within the patch are unaliased from the same acquired data[20]. Following these steps, the noise component of a given $\mathrm{M} \times \mathrm{N}$ Casorati matrix $\mathbf{Y}$ has zero-mean i.i.d. entries, and the threshold that ensures all components that are indistinguishable from Gaussian noise are removed is given as the largest singular value of the noise matrix. While this threshold is asymptotically specified by the Marchenko-Pastur distribution, for practical finite matrix sizes, we numerically estimate this value via a Monte-Carlo simulation[20] utilizing the variance of the measured thermal noise in k-space, σ_{ksp}^2. In our implementation, M is chosen as 11×N, which was empirically determined in our prior study[20]. The thresholded Casorati matrices are then combined using patch averaging[48]. Finally, the denoised volumes $\boldsymbol{m}^{\boldsymbol{NORDIC}}(\mathbf{r},\mathrm{t})$ are multiplied back with the g-factor map $g(\mathbf{r})$ to correct the signal intensities.

Participants

Two (1 male and 1 female) healthy right-handed subjects (age: 32 and 33) participated in the study. Both subjects had normal, or corrected vision and provided written informed consent. The local IRB at the University of Minnesota approved the experiments.

Stimuli and procedure

The experimental procedure consisted of a standard 12 seconds on 12 seconds off block design paradigm (se Figure 1A). The stimuli consisted of a center (i.e. target) and a surround square checkerboard counterphase flickering (at 6 Hz) gratings (Figure 1A) subtending approximately 6.5 degrees of visual angle. Stimuli were centered on a background of average luminance (25.4 cd/m^2, 23.5-30.1). Stimuli were presented on a Cambridge Research Systems BOLDscreen 32 LCD monitor positioned at the head of the 7T scanner bed (resolution 1920, 1080 at 120 Hz; viewing distance ~89.5 cm.) using Mac Pro computer. Stimulus presentation was

controlled using Psychophysics Toolbox (3.0.15) based codes. Participants viewed the images though a mirror placed in the head coil.

Each run lasted just over two and a half minutes (i.e. 118 volumes at 1.350 seconds TR), beginning and ending with a 12 seconds red fixation dot centered on a gray background. Within each run, each visual condition, target and surround, was presented 3 times. Participants underwent 8 runs each, 2 of which were used to compute the region of interest and excluded from subsequent analyses. Participants were instructed to minimize movement and keep fixation locked on the center fixation dot throughout the experimental runs.

MR Imaging Acquisition and Processing

All functional MRI data were collected with a 7T Siemens Magnetom System with a single transmit and 32-channel receive NOVA head coil. T2*-weighted images were collected using sequence parameters (TR 1.350s, Multiband factor 2, 3-fold in-plane undersampling along the phase encode direction, 6/8ths Partial Fourier, 0.8mm isotropic nominal resolution, TE 26.4ms, Flip Angle 58.2°, Bandwidth 1190Hz). The flip angle was optimized to maximize the signal across the brain for the given TR. For each participant, shimming to improve B0 homogeneity over occipital regions was conducted manually.

T1-weighted anatomical images were obtained on a 3T Siemens Magnetom Prisma^{fit} system using an MPRAGE sequence (192 slices; TR, 1900 ms; FOV, 256 x 256 mm; flip angle 9°; TE, 2.52 ms; 0.8 mm isotropic voxels) . Anatomical images were used for visualization purposes and to define the cortical grey matter ribbon. This was done in BrainVoyager via automatic segmentation based on T1 intensity values and subsequent manual corrections. All analyses were subsequently confined within the gray matter.

Functional data Preprocessing

Functional data preprocessing was performed in BrainVoyager. Preprocessing was kept at a minimum and constant across reconstructions. Specifically, we performed slice scan timing corrections (sinc interpolation), 3D rigid body motion correction (sinc interpolation), where all volumes for all runs were motion corrected relative to the first volume of the first run acquired, and low drift removals (i.e. temporal high pass filtering) using a GLM approach with a design matrix continuing up to the 2nd order discrete cosine transform basis set. No spatial nor temporal smoothing was applied. Functional data were aligned to anatomical data with manual adjustments and iterative optimizations. Anatomical alignment was performed on the Single Band Reference

(SBRef) image which was acquired to calibrate coil sensitively profiles prior to the multiband acquisition and has no slice acceleration or T_1-saturation, yielding higher contrast[49].

GLMs and tSNR

All subsequent analyses were performed in Matlab using a set of tools developed inhouse.

Temporal signal-to-noise ratio (tSNR) was computed by dividing the mean (over time) of the detrended time-courses by its standard deviation independently per voxel, run and subject.

To quantify the extent of stimulus evoked activation, we performed general linear model (GLM) estimation (with ordinary least squares minimization). Design matrices (DMs) were generated by convolution of a double gamma function with a "boxcar" function (representing onset and offset of the stimuli). We computed both single trial as well as condition-based GLMs. The latter, where DMs had one predictor per condition, were used to assess the differences in extent and magnitude in activation between NORDIC and Standard images. The former, where the DMs had one predictor per trial per condition, produced single trials activation estimates that were used to assess the stability (see "*NORDIC vs. Standard statistical analyses"* paragraph below) of the responses evoked by the target condition for each voxel within the left retinotopic representation of the target in V1 (see below).

ROIs definition

Out of the 8 recorded runs, 2 runs (identical for each reconstruction type) were used to define a region of interest (ROI). Specifically, we performed a classic GLM on 4 concatenated runs (2 reconstructed with NORDIC and 2 with the standard algorithm) and computed the differential map by contrasting the t-values elicited by the target to that elicited by the surround. GLM t-values, can be thought of beta estimates divided by GLM standard error according to this equation:

$$t = c'b/\sqrt{Var(e)c'(X'X)^{-1}c} \qquad \text{(eq. 2)}$$

where b represents the beta weights, c is a vector of 1, -1 and 0 indicating the conditions to be contrasted, e is the GLM residuals and X the design matrix. We then thresholded this map ($p<.05$ Bonferroni corrected) to define the left hemisphere retinotopic representation of the target stimulus within the grey matter boundaries. This procedure was implemented to provide an identical ROI across reconstruction types, however, it resulted in effectively doubling the number of data points available, which could not be treated as independent anymore. To partially account

for this, we adjusted the GLM degrees of freedom used to compute the t-maps to be equal to those of 2 rather than 4 runs.

GLMs for experimental runs

Independently per reconstruction type, for the condition-based scenario, GLMs were performed for each single run as well as for multiple runs (i.e. concatenating 2 or more experimental runs and design matrices to estimate BOLD responses). For the multiple run scenarios, we estimated the percent signal change beta weights and the related t-values for 2, 3, 4, 5 and 6 runs. For each n-runs GLM, we computed independent GLMs for all possible run combinations (see the "*Comparing extent of activation*" paragraphs for more details).

NORDIC vs. Standard statistical analyses

In order to evaluate the impact of NORDIC denoising on BOLD based GE-EPI fMRI images, the following analyses were performed. Standard tSNR was computed as previously described. To assess statistically significant differences in average tSNR across reconstruction types, we first computed the mean tSNR (using the 20% trimmed mean, which is more robust to extreme values,[24]) across all voxels in the brain for each of the 8 runs. We then carried out 2-tailed paired sample t-tests between average tSNRs for NORDIC and Standard images across all runs.

Moreover, to test for statistical significant differences in stimulus-evoked BOLD amplitudes and noise levels across reconstruction algorithms, we compared the ROI voxel mean percent signal change beta estimates and related t-values elicited by the target condition independently per subject. We used the 18 responses elicited by the 3 stimulus presentations within each of the 6 runs. To account for the fact that trials within each run are not independent, while the runs are, we implemented a Linear Mixed-Effect Model in Matlab (The Mathworks Inc, 2014) according to the equation:

$$\text{Data} \sim \text{Cond} + (1|\text{runs}) + (1|\text{trials}) \qquad \text{(eq. 3)}$$

Linear Mixed-Effect Model allows estimating fixed and random effects, thus allowing modeling variance dependencies within terms. Model coefficients were estimated by means of maximum likelihood estimation.

To assess differences in the precision of BOLD PSC estimates across reconstructions, we computed the cross-validates R2 for single runs GLMs. This was achieved by deriving the beta weights using a given "training" run, and testing how well these estimates predicted single voxel activation for all other "test" runs. Single voxel cross validated R2 (also known as coefficient of determination) was computed according to the equation:

$$R2 = 1 - SS_{error} / SS_{total} \quad \text{(eq. 4)}$$

and, in our specific case

$$SS_{error} = \Sigma\, (f(x) / x)^2 \quad \text{(eq. 5)}$$

$$SS_{total} = \Sigma\, (x_{(i)} - \mu(x))^2 \quad \text{(eq. 6)}$$

In eq. 5 and 6, x is the empirical time-course of the test run, $x_{(i)}$ represents the ith point of the empirical time-course of the test run x, $\mu(x)$ is the time-course mean, and $f(x)$ is the predicted time course computed by multiplying the design matrix of a given training run by the beta estimates derived on a different test run.

We computed all possible unique combinations of training the model on a given run and testing on all remaining runs, leading to 15 R2s per voxel. To infer statistical significance, we carried out paired sample t-tests across the 15 cross-validated R2 (averaged across all ROI voxels) for NORDIC and Standard images.

To assess the stability and thus the reliability of single trials response estimates we computed the standard deviation across percent signal change amplitudes elicited by each single presentation (i.e. single trial) of the target stimulus for every run, voxel and reconstruction type. To infer statistical significance between these stability estimates for NORDIC and Standard, independently per subject we carried out 2 tests: 1) we performed 2-tailed paired sample t-tests across runs; 2) we computed 95% bootstrap confidence intervals as follows. First, for a given subject, we computed the difference between the single trials' standard deviations of NORDIC and Standard data. For each bootstrap iteration, we then sampled with replacement the runs, computed the mean across the sampled runs and stored the value. We repeated this operation 10000 times, leading to 10000 means. We sorted these 10000 means and selected the 97.5 and the 2.5 percentiles (representing the 95% bootstrapped confidence intervals of the difference). Statistical significance was inferred when 95% bootstrap confidence interval did not overlap with 0.

Comparing extent of activation

We further compared the extent of activation across reconstructions for the GLMs computed on 1 and multiple runs by quantifying the number of active voxels at a fixed t-value threshold. To this end, we computed the t-map for the contrast target > surround. For each GLM, we then counted the number of significant voxels at $t \geq 5.7$ (corresponding to $p<.05$ Bonferroni corrected for the Standard images) within the ROI. As we intended to understand and quantify the difference in extent of activation between NORDIC and Standard reconstructions, we compared GLM computed on 1 NORDIC run versus 1, 2, 3, 4, 5 and 6 runs of Standard GLMs. To ensure that any potential difference was not related to run-to-run variance, we implemented the following procedure. Firstly, we computed GLMs for all possible unique run combinations. This led to six data points for single run GLMs (i.e. 1 GLM per experimental run), 15 data points for GLMs computed on 2 concatenated runs (e.g. runs 1-2; 1-3;1-4;1-5;1-6; 2-3 etc.), 20 data points for GLMs computed on 3 concatenated runs; 15 data points for GLMs computed on 4 concatenated runs; 6 data points for GLMs computed on 5 concatenated runs and 1 data point for the GLM computed on 6 concatenated runs. For each run combination, we counted the significant number of active voxels at our statistical threshold and stored those numbers. Within each n-run GLM (where n represents the number of concatenated runs), we then proceeded to compute 95% bootstrap confidence interval on the mean of the active number of voxels across all possible run combinations. This was achieved by sampling with replacement the number of significantly active voxels estimated for each combination of runs and computing the mean across the bootstrap sample. We repeated this operation 1000 times to construct a bootstrap distribution and derive 95% bootstrap confidence interval[24]. This procedure not only ensured sampling from all runs, but it also decreased the impact of extreme values[24].

Quantifying BOLD images Smoothness

Global smoothness estimates from each reconstruction prior to preprocessing ('pre') and data following all preprocessing, just prior to the GLM ('post'). This was performed using 3dFWHMx from AFNI[25] using the '-ACF' command. The data were detrended using the default settings from 3dFWHMx with the '-detrend' command. As we are interested in the smoothness within the brain, we also used the '-automask' command in order to generate an intensity-based brain mask, based on the median value of each run. This method iterates through various background clipping parameters to generate a contiguous brain only volume, that excludes the

external areas of low signal. The spatial autocorrelation is estimated from the data using a Gaussian plus mono-exponential model, which accounts for possible long-tail spatial autocorrelations found in fMRI data. This estimated FWHM, in mm, from this fitted autocorrelation function is used as an estimate of the smoothness of the data. This estimate was derived for all of the runs, excluding the held-out runs used for ROI creation. For each subject, smoothness was averaged within each stage across the 6 experimental runs to evaluate if global smoothness was markedly increased due to the reconstruction method. Paired sample t-tests were carried out between estimated FWHM parameters for the NORDIC and Standard reconstructions to infer statistical significance.

Functional point spread function

Functional point spread function (PSF) was computed according to[23]. In brief we estimated the BOLD functional PSF on the GLM computed on all runs concatenated. This analysis was performed independently per run and separately for the Standard and NORDIC reconstructions. Within each run, we first identified the anterior most retinotopic representation of the target's edge in V1. This was achieved by computing the contrast target > surround and identifying those voxels showing differential BOLD closest of 0 (Figure 6). Then, using BrainVoyager, we flattened this portion of the cortex to produce Laplace-based equipotential grid-lines in the middle of the cortical ribbon. To increase the precision of the PSF measurement, we upsampled the BOLD activation maps to 0.1 mm isotropic voxel. We drew 10 traces orthogonal to the retinotopically anterior most edge of the target. We thus superimposed these traces to the activity elicited by the target condition and, from the target's edge, we measured the slope of BOLD amplitude decrease along the traces. PSF was quantified by fitting a model to the mean of the 10 traces with 3 free parameters consisting of a step-function like (representing infinitely precise PSF) convolved with a gaussian[23]. The 3 parameters were the width of the gaussian (representing functional precision – see[23], the retinotopic location of the edge and a multiplicative constant. Parameter fitting was performed in Matlab using the *lsqcurvefit* function, with sum of squares as stress metric. Paired sample t-tests across the 8 runs were then carried out between the Gaussian widths for NORDIC and Standard images to infer statistical significance.

ACKNOWLEDGEMENTS: The authors thank Prof. Kendrick Kay, University of Minnesota, for helpful discussion and comments. This work was supported by NIH grants U01EB025144 (K.U.), P41 EB027061 (K.U.), P30 NS076408 (K.U.) and RF1 MH116978 (E.Y.), and RF1 MH117015 (Geoffrey Ghose, University of Minnesota).

AUTHOR CONTRIBUTIONS: LV, EY, KU designed the experiments. LV and LD acquired data. LD analyzed the 3T data; FdM analyzed the auditory data; LV carried out all remaining analyses. SM and MA conceptualized the denoising methodology, SM wrote the denoising algorithm and executed it on the acquired data. LV, KU, SM, MA wrote the paper. All authors discussed the results, the manuscript, and edited the manuscript.

COMPETING INTEREST: The authors confirm that there are no competing interests.

REFERENCES

1. Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H. & Ugurbil, K. *Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging*. (1992) Proc Natl Acad Sci U S A **89**, 5951-5955.
2. Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., Cheng, H.-M., Brady, T. J. & Rosen, B. R. *Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation*. (1992) Proc Natl Acad Sci U S A **89**, 5675-5679.
3. Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S. & Hyde, J. S. *Time course EPI of human brain function during task activation*. (1992) Magn Reson Med **25**, 390-397.
4. Van Essen, D. C. & Ugurbil, K. *The future of the human connectome*. (2012) Neuroimage **62**, 1299-1310.
5. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T. & Collins, R. *UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age*. (2015) PLoS Med **12**, e1001779.
6. Jorgenson, L. A., Newsome, W. T., Anderson, D. J., Bargmann, C. I., Brown, E. N., Deisseroth, K., Donoghue, J. P., Hudson, K. L., Ling, G. S., MacLeish, P. R., Marder, E., Normann, R. A., Sanes, J. R., Schnitzer, M. J., Sejnowski, T. J., Tank, D. W., Tsien, R. Y., Ugurbil, K. & Wingfield, J. C. *The BRAIN Initiative: developing technology to catalyse neuroscience discovery*. (2015) Philos Trans R Soc Lond B Biol Sci **370**.
7. Ugurbil, K. *Imaging at ultrahigh magnetic fields: History, challenges, and solutions*. (2018) Neuroimage **168**, 7-32.
8. Ugurbil, K. *What is feasible with imaging human brain function and connectivity using functional magnetic resonance imaging*. (2016) Philos Trans R Soc Lond B Biol Sci **371**.
9. De Martino, F., Yacoub, E., Kemper, V., Moerel, M., Uludag, K., De Weerd, P., Ugurbil, K., Goebel, R. & Formisano, E. *The impact of ultra-high field MRI on cognitive and computational neuroimaging*. (2018) Neuroimage **168**, 366-382.

10. Dumoulin, S. O., Fracasso, A., van der Zwaag, W., Siero, J. C. W. & Petridou, N. *Ultra-high field MRI: Advancing systems neuroscience towards mesoscopic human brain function*. (2018) Neuroimage **168**, 345-357.
11. He, X., Erturk, M. A., Grant, A., Wu, X., Lagore, R. L., DelaBarre, L., Eryaman, Y., Adriany, G., Auerbach, E. J., Van de Moortele, P. F., Ugurbil, K. & Metzger, G. J. *First in-vivo human imaging at 10.5T: Imaging the body at 447 MHz*. (2020) Magn Reson Med **84**, 289-303.
12. Sadeghi-Tarakameh, A., DelaBarre, L., Lagore, R. L., Torrado-Carvajal, A., Wu, X., Grant, A., Adriany, G., Metzger, G. J., Van de Moortele, P. F., Ugurbil, K., Atalar, E. & Eryaman, Y. *In vivo human head MRI at 10.5T: A radiofrequency safety study and preliminary imaging results*. (2020) Magn Reson Med **84**, 484-496.
13. Nowogrodzki, A. *The world's strongest MRI machines are pushing human imaging to new limits*. (2018) Nature **563**, 24-26.
14. Hoult, D. I. & Richards, R. E. *The signal-to-noise ratio of the nuclear magnetic resonance phenomenon*. (1976) J. Magn. Reson. **24**, 71 - 85.
15. Edelstein, W. A., Glover, G. H., Hardy, C. J. & Redington, R. W. *The intrinsic signal-to-noise ratio in NMR imaging*. (1986) Magn Reson Med **3**, 604-618.
16. Pruessmann, K. P., Weiger, M., Scheidegger, M. B. & Boesiger, P. *SENSE: sensitivity encoding for fast MRI*. (1999) Magn Reson Med **42**, 952-962.
17. Triantafyllou, C., Hoge, R. D., Krueger, G., Wiggins, C. J., Potthast, A., Wiggins, G. C. & Wald, L. L. *Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters*. (2005) Neuroimage **26**, 243-250.
18. Bianciardi, M., Fukunaga, M., van Gelderen, P., Horovitz, S. G., de Zwart, J. A., Shmueli, K. & Duyn, J. H. *Sources of functional magnetic resonance imaging signal fluctuations in the human brain at rest: a 7 T study*. (2009) Magn Reson Imaging **27**, 1019-1029.
19. Kay, K. N., Rokem, A., Winawer, J., Dougherty, R. F. & Wandell, B. A. *GLMdenoise: a fast, automated technique for denoising task-based fMRI data*. (2013) Front Neurosci **7**, 247.
20. Moeller, S., Pishardy, P. K., Ramanna, S., Lenglet, C., Wu, X., Yacoub, E., Ugurbil, K. & Akcakaya, M. *NOise Reduction with DIstribution Corrected (NORDIC) PCA in dMRI with complex-valued parameter-free locally low-rank processing*. (2020) bioRxiv *Published on line before Print; DOI: 10.1101/2020.08.25.267062*, 2020.2008.2025.267062.
21. Ugurbil, K., Xu, J., Auerbach, E. J., Moeller, S., Vu, A. T., Duarte-Carvajalino, J. M., Lenglet, C., Wu, X., Schmitter, S., Van de Moortele, P. F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D. A., Smith, S. M., Miller, K. L., Sotiropoulos, S. N., Jbabdi, S., Andersson, J. L., Behrens, T. E., Glasser, M. F., Van Essen, D. C., Yacoub, E. & Consortium, W. U.-M. H. *Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project*. (2013) Neuroimage **80**, 80-104.
22. Barth, M., Breuer, F., Koopmans, P. J., Norris, D. G. & Poser, B. A. *Simultaneous multislice (SMS) imaging techniques*. (2016) Magn Reson Med **75**, 63-81.
23. Shmuel, A., Yacoub, E., Chaimow, D., Logothetis, N. K. & Ugurbil, K. *Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 Tesla*. (2007) Neuroimage **35**, 539-552.
24. Wilcox, R. R. *Introduction to Robust Estimation of Hypothesisi Testing*. *Published on line before Print; DOI*, (Elsevier Academic Press, 2005).
25. Cox, R. W. *AFNI: software for analysis and visualization of functional magnetic resonance neuroimages*. (1996) Comput Biomed Res **29**, 162-173.

26. Andersson, J. L. & Sotiropoulos, S. N. *Non-parametric representation and prediction of single- and multi-shell diffusion-weighted MRI data using Gaussian processes*. (2015) Neuroimage **122**, 166-176.
27. Manjon, J. V., Coupe, P., Concha, L., Buades, A., Collins, D. L. & Robles, M. *Diffusion weighted image denoising using overcomplete local PCA*. (2013) PLoS One **8**, e73021.
28. Veraart, J., Novikov, D. S., Christiaens, D., Ades-Aron, B., Sijbers, J. & Fieremans, E. *Denoising of diffusion MRI using random matrix theory*. (2016) Neuroimage **142**, 394-406.
29. Vaidya, M. V., Sodickson, D. K. & Lattanzi, R. *Approaching Ultimate Intrinsic SNR in a Uniform Spherical Sample with Finite Arrays of Loop Coils*. (2014) Concepts Magn Reson Part B Magn Reson Eng **44**, 53-65.
30. Hu, X. & Kim, S. G. *Reduction of signal fluctuation in functional MRI using navigator echoes*. (1994) Magn Reson Med **31**, 495-503.
31. Lund, T. E., Madsen, K. H., Sidaros, K., Luo, W. L. & Nichols, T. E. *Non-white noise in fMRI: does modelling have an impact?* (2006) Neuroimage **29**, 54-66.
32. Xu, J., Moeller, S., Auerbach, E. J., Strupp, J., Smith, S. M., Feinberg, D. A., Yacoub, E. & Ugurbil, K. *Evaluation of slice accelerations using multiband echo planar imaging at 3 T*. (2013) Neuroimage **83**, 991-1001.
33. 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. *Correspondence of the brain's functional architecture during activation and rest*. (2009) Proc Natl Acad Sci U S A **106**, 13040-13045.
34. Bianciardi, M., van Gelderen, P., Duyn, J. H., Fukunaga, M. & de Zwart, J. A. *Making the most of fMRI at 7 T by suppressing spontaneous signal fluctuations*. (2009) Neuroimage **44**, 448-454.
35. Fox, M. D., Snyder, A. Z., Zacks, J. M. & Raichle, M. E. *Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses*. (2006) Nature neuroscience **9**, 23-25.
36. Kruger, G. & Glover, G. H. *Physiological noise in oxygenation-sensitive magnetic resonance imaging*. (2001) Magn Reson Med **46**, 631-637.
37. Kruger, G., Kastrup, A. & Glover, G. H. *Neuroimaging at 1.5 T and 3.0 T: comparison of oxygenation-sensitive magnetic resonance imaging*. (2001) Magn Reson Med **45**, 595-604.
38. Hyde, J. S., Biswal, B. B. & Jesmanowicz, A. *High-resolution fMRI using multislice partial k-space GR-EPI with cubic voxels*. (2001) Magn Reson Med **46**, 114-125.
39. Yacoub, E., Van De Moortele, P. F., Shmuel, A. & Ugurbil, K. *Signal and noise characteristics of Hahn SE and GE BOLD fMRI at 7 T in humans*. (2005) Neuroimage **24**, 738-750.
40. Triantafyllou, C., Polimeni, J. R. & Wald, L. L. *Physiological noise and signal-to-noise ratio in fMRI with multi-channel array coils*. (2011) Neuroimage **55**, 597-606.
41. Chen, J. E., Polimeni, J. R., Bollmann, S. & Glover, G. H. *On the analysis of rapidly sampled fMRI data*. (2019) Neuroimage **188**, 807-820.
42. Triantafyllou, C., Hoge, R. D. & Wald, L. L. *Effect of spatial smoothing on physiological noise in high-resolution fMRI*. (2006) Neuroimage **32**, 551-557.
43. Chaimow, D., Yacoub, E., Ugurbil, K. & Shmuel, A. *Spatial specificity of the functional MRI blood oxygenation response relative to neuronal activity*. (2018) Neuroimage **164**, 32-47.
44. Sotiropoulos, S. N., Moeller, S., Jbabdi, S., Xu, J., Andersson, J. L., Auerbach, E. J., Yacoub, E., Feinberg, D., Setsompop, K., Wald, L. L., Behrens, T. E., Ugurbil, K. & Lenglet, C. *Effects of image reconstruction on fiber orientation mapping from*

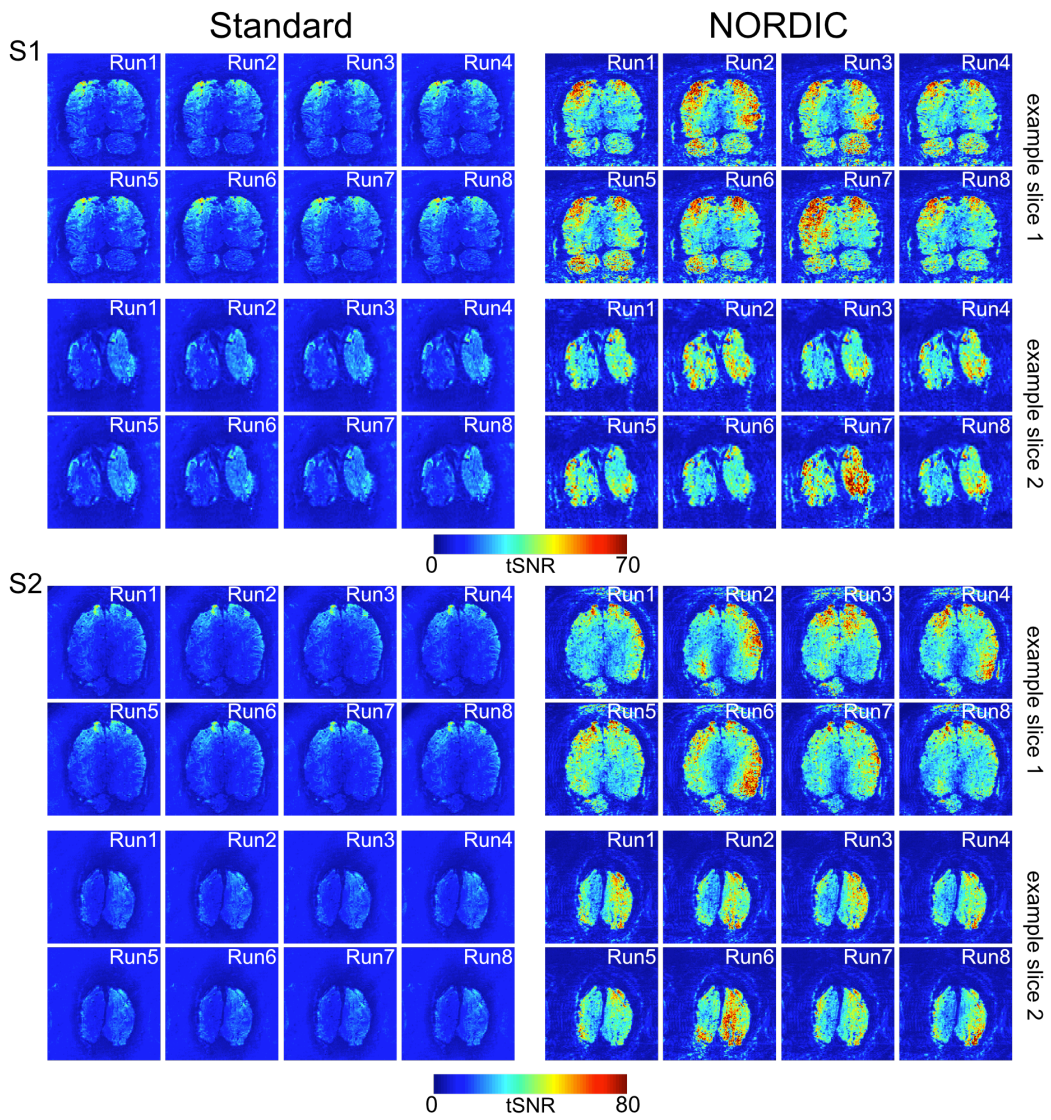
multichannel diffusion MRI: reducing the noise floor using SENSE. (2013) Magn Reson Med **70**, 1682-1689.

45. Moeller, S., Pisharady, P. K., Andersson, J., Akcakaya, M., Harel, N., Ma, R. E., Wu, X., Yacoub, E., Lenglet, C. & Ugurbil, K. *Diffusion imaging in the post HCP era*. (2020) Journal of Magnetic Resonance Imaging *Published on line before Print; DOI: 10.1002/jmri.27247*.
46. Aja-Fernández, S. & Vegas-Sánchez-Ferrero, G. *Statistical analysis of noise in MRI*. *Published on line before Print; DOI: 10.1007/978-3-319-39934-8* (Springer Berlin Heidelberg, 2016).
47. Breuer, F. A., Kannengiesser, S. A., Blaimer, M., Seiberlich, N., Jakob, P. M. & Griswold, M. A. *General formulation for quantitative G-factor calculation in GRAPPA reconstructions*. (2009) Magn Reson Med **62**, 739-746.
48. Katkovnik Vladimir, Alessandro Foi, Karen Egiazarian & Astola, J. *From local kernel to nonlocal multiple-model image denoising.* (2010) International journal of computer vision *Published on line before Print; DOI*.
49. Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D. A., Griffanti, L., Harms, M. P., Kelly, M., Laumann, T., Miller, K. L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A. Z., Vu, A. T., Woolrich, M. W., Xu, J., Yacoub, E., Ugurbil, K., Van Essen, D. C., Glasser, M. F. & for the WU-Minn HCP Consortium. *Resting-state fMRI in the Human Connectome Project*. (2013) Neuroimage **80**, 144-168.

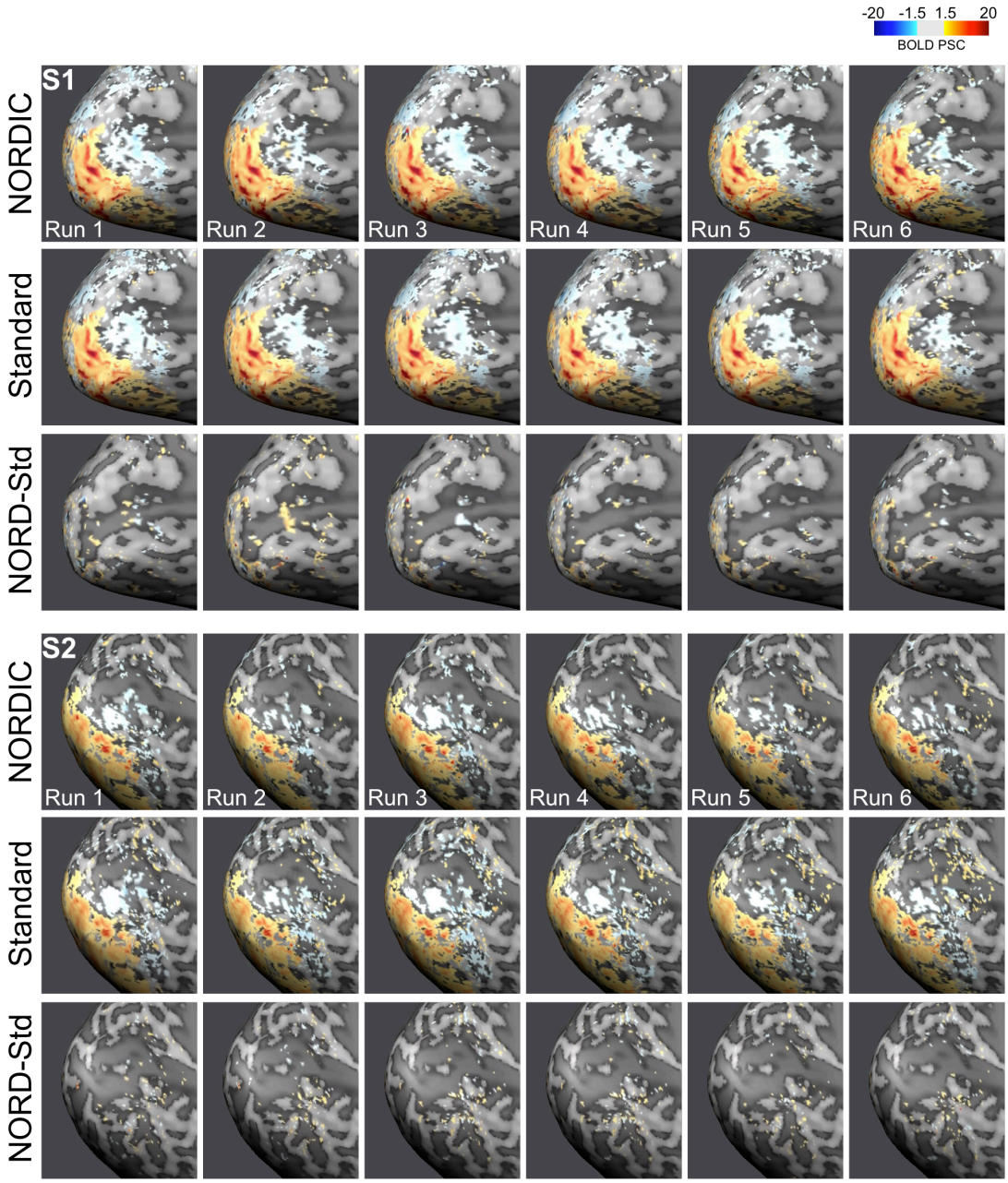
Supplementary Material

The supplemental information presented in this section falls into two categories**: I)** Supplementary Figs. 1 through 4, showing additional analyses and results based on the 7 Tesla data presented in the main body of the paper to supplement the conclusions reached from these 7T data; **II)** additional data sets and discussion demonstrating the wide applicability of NORDIC across field strengths, cortical regions, stimulation and/or task paradigms, and acquisition strategies.

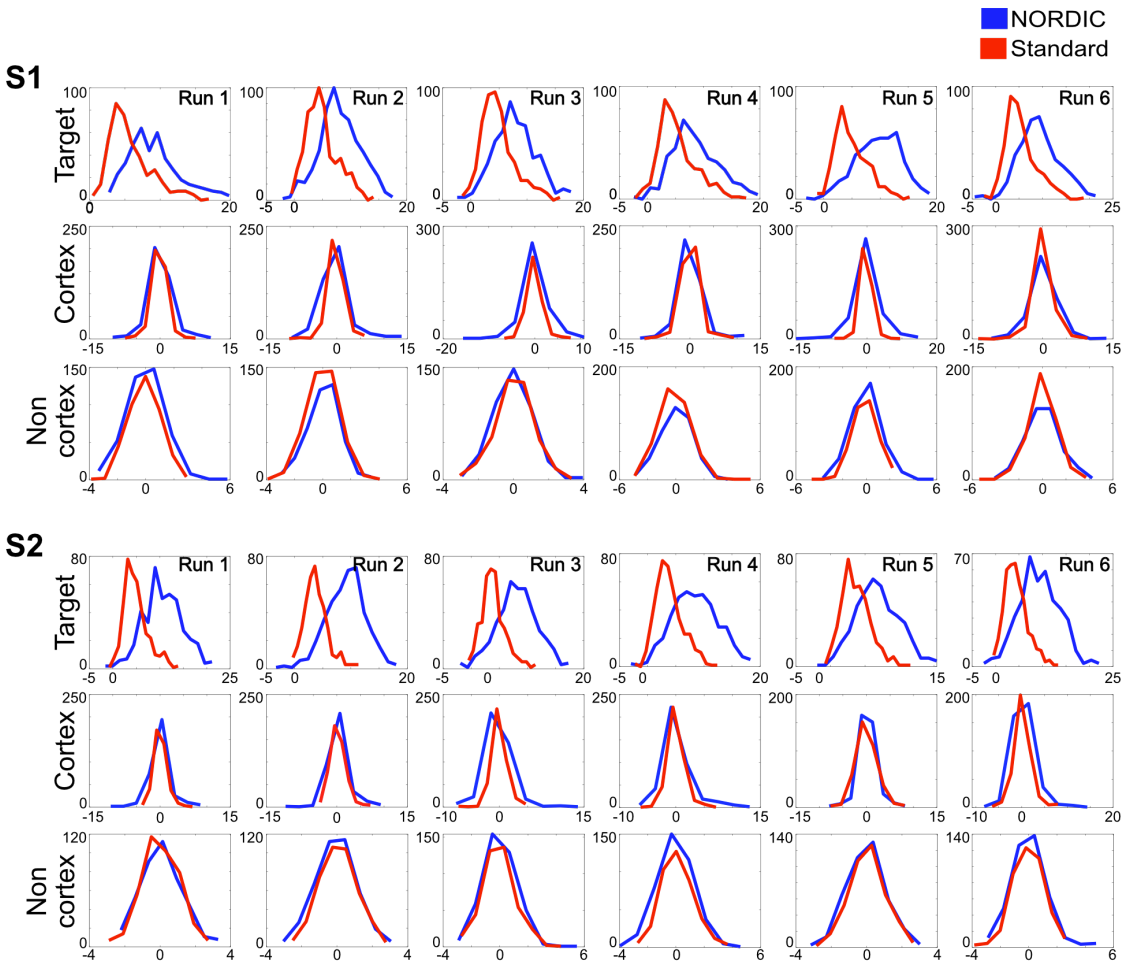
I. Supplementary Figures based on the 7T data presented in the main body of the paper.



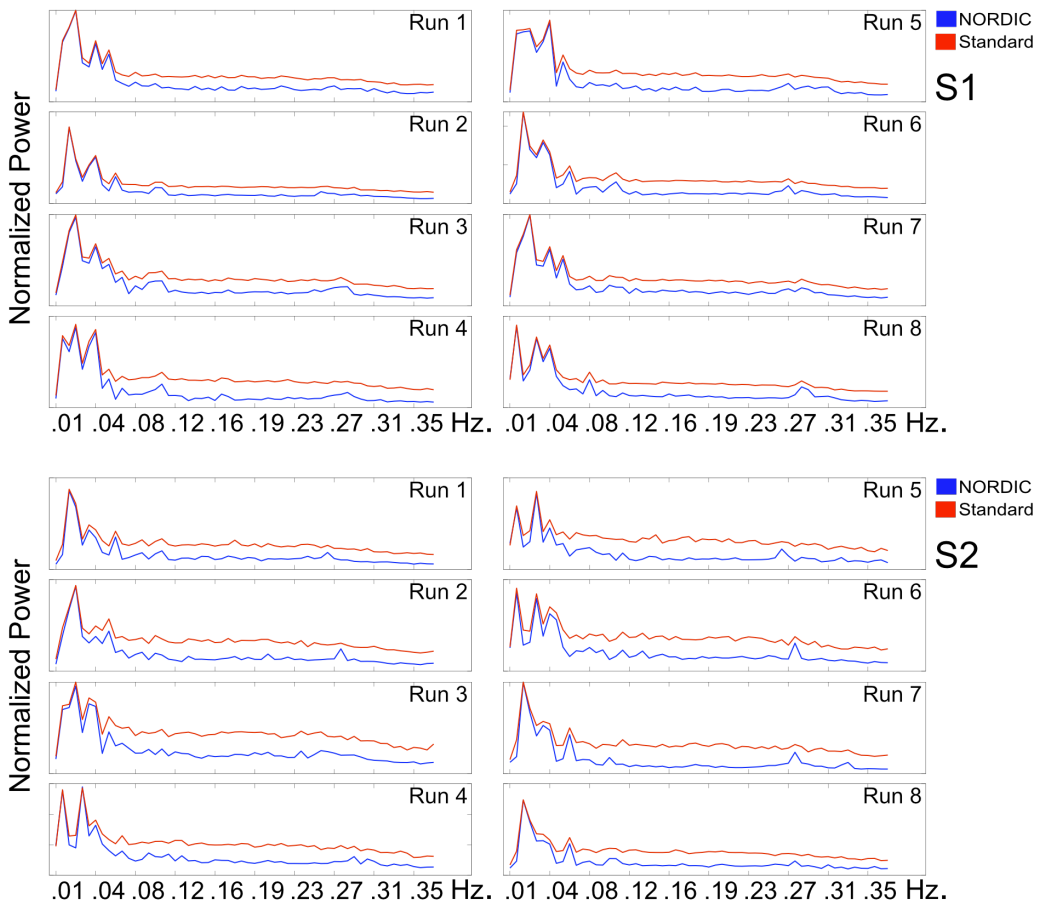
Supplementary Figure 1. Temporal SNR (tSNR) maps of 2 exemplar slices in both subjects for all 8 different runs for Standard (left) and NORDIC (right) reconstructions. Slice 1, represents one of the anterior most slices in the covered volume; while slice 2 is an occipital slice that includes a portion of V1.



Supplementary Figure 2. Single run percent BOLD signal change (PSC) maps for both subjects. For each subject (S1 and S2) the top row portrays the NORDIC single run BOLD PSC maps elicited by the target condition. The second row is equivalent to the first, but for Standard images. The bottom row shows the BOLD difference between NORDIC and Standard. As evident in this figure, BOLD PSC maps are highly comparable across reconstructions.



Supplementary Figure 3. NORDIC (blue) and Standard (red) T-values distribution for the contrast Target > 0 (see methods) for 3 ROIs. In all plots, the x axis depicts the magnitude of the t-values, while the y axis shows the voxel count. For each subject (S1 and S2) the top row shows the histogram of the t-values for all the voxels within the target ROI. The second row shows the histogram of the t-values for a number of randomly selected voxels within the cortex not belonging to the target ROI. The third row shows the histogram of the t-values for a number of randomly selected voxels outside the cortex. For the cortex and non-cortex ROIs the number of voxels was chosen to match that of the target ROI. The t-values distribution for NORDIC and Standard are highly non-overlapping for the target voxels. In light of the comparable PSC amplitudes across reconstructions, this set of figures indicates a dramatic noise reduction for NORDIC images. The histograms for voxels within the cortex but outside the target ROI indicate a lesser improvement in data quality compared to the target voxels; while the histograms for non-cortical voxels are highly overlapping. This figure highlights how NORDIC denoising selectively improves data quality by removing thermal noise in relevant portions of the cortex (i.e. those where BOLD activation is expected).



Supplementary Figure 4. Fast Fourier Transform (FFT) for all subjects (S1 and S2) and runs for NORDIC (blue) and Standard (red) images. FFT was computed independently per voxel within the target ROI. The magnitude output of the FFTs were then averaged across all target voxels to produce the above power spectra. The x axis portrays frequencies in Hertz. For NORDIC (but not Standard) power spectra, a peak at approximately .27 Hz – representing the respiratory frequency - is visible for all subjects and runs. As respiration belongs to physiological noise, these figures not only confirm that NORDIC denoising operates in the thermal noise regime, but also that structured physiological noise such as respiration becomes more easily detectable after thermal noise removal. Lower frequencies, representative of neuronal responses, remain largely unaffected, indicating that NORDIC denoising does not compromise neural BOLD activity.

II. SUPLEMENTARY DATA and DISCUSSION DEMONSTRATING WIDE APPLICABILITY of NORDIC, across field strengths, spatial resolutions, cortical regions, stimulation and/or task paradigms, and acquisition strategies.

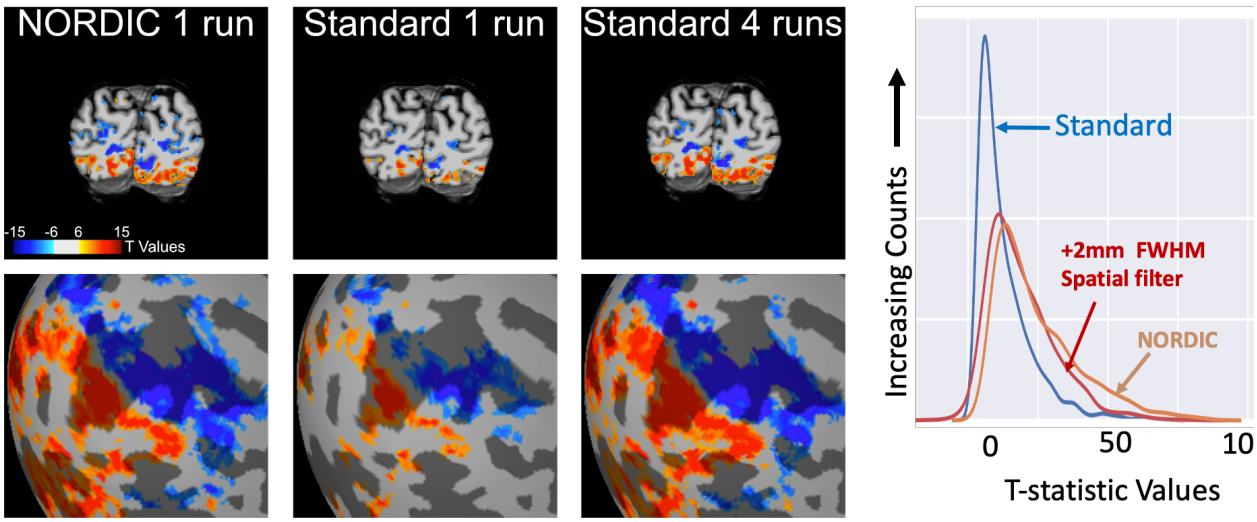
i) NORDIC applications at 3 Tesla:

In the manuscript, we presented only high resolution 7 Tesla fMRI data. However, vast majority of the functional imaging studies are carried out at 3 Tesla using supra-millimeter resolutions. Therefore, one of the most important indicators for the wide applicability of the NORDIC method would be to demonstrate that it substantially improves such 3 Tesla data.

A ”gold standard” of fMRI acquisition at 3 Tesla is the approach employed by the Human Connectome Project (HCP), which uses 2 mm isotropic resolution and Simultaneous Multi Slice (SMS)/Multiband (MB) acquisitions [1,2]. In order to evaluate the impact of NORDIC on 3T fMRI data, we employed exactly this HCP acquisition approach to obtain functional maps at 3T for the same target and surround visual stimulation paradigm employed and described in the main body (see Fig. 1A in the Manuscript). Briefly, this is a ~2.5 min long paradigm within which retinotopically arranged “target” and “surround” stimuli are presented 3 times each for 12 s followed by 12 s blank epochs; such a ~2.5 fMRI “run” was then repeated multiple times.

The results obtained with this paradigm at 3T using the HCP acquisition protocol are illustrated in Supplementary Fig. 5. The use of 4 concatenated runs representing ~10 mins of data are needed to generate an fMRI map (t-statistics map with thresholding) equivalent to that obtained from a single run (~ 2.5 min of data) using NORDIC denoising. This improvement is expected based on the distribution of t-values for NORDIC *versus* Standard reconstruction in these data (Supplementary Fig. 5, right most panel).

The t-statistics are also displayed in Supplementary Fig. 5 (right-most panel) when the fMRI data was spatially smoothed with a 2 mm Full Width Half Maximum (FWHM) filter. Spatial filtering is

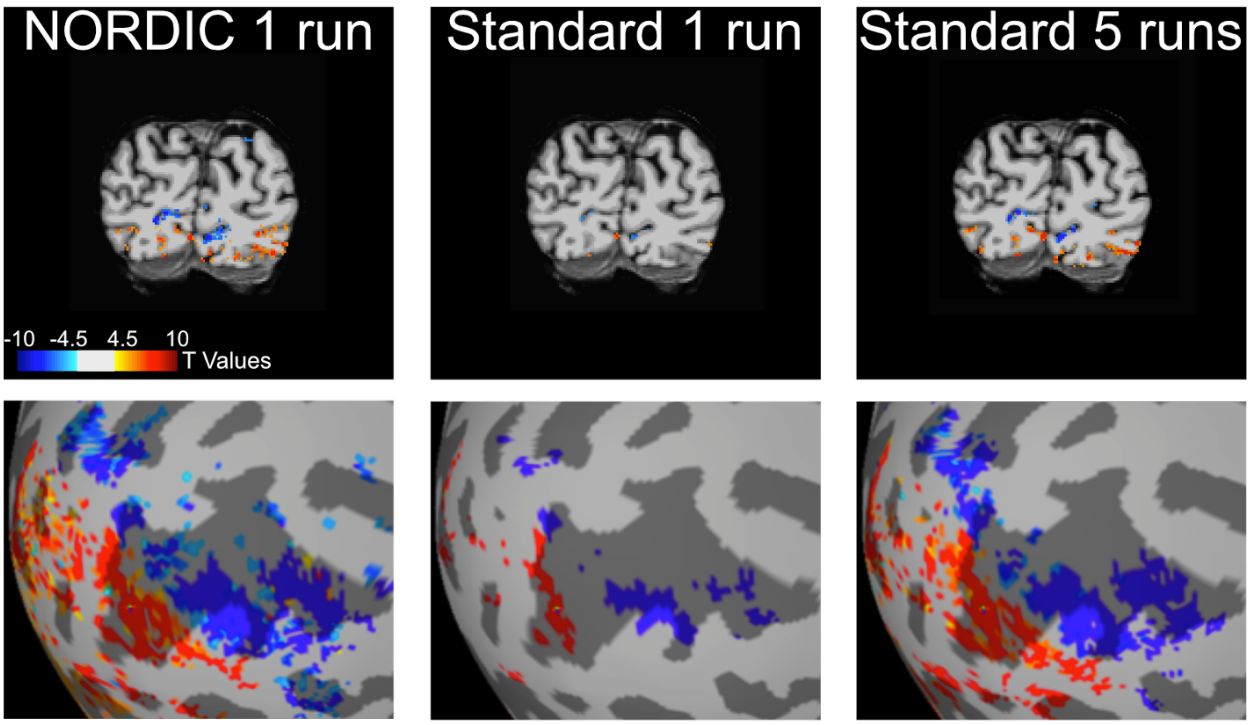


Supplementary Figure 5: The effect of NORDIC denoising on an fMRI study carried out at 3 Tesla using the standard acquisition parameters of the Human Connectome Project (HCP): Spatial resolution= 2 mm; TR= 800 ms; Simultaneous Multislice(SMS)/Multiband (MB) EPI acquisition with MB factor of 8. Stimulus employed was identical to that implemented in the main study described in the manuscript (see Methods), where flickering gratings for target or surround were presented in 12 second blocks interleaved by 12 second fixation periods. The figure shows t-maps with t value ≥|6| for the contrast target (red) > surround (blue) in volume space for a representative slice in the visual cortex (top row), and in the inflated cortical surface (bottom row). T values >|15| appear as the same color as t=|15|. Approximately four runs with standard reconstruction (~10 min of data) are required to achieve the extent of activation comparable a single NORDIC run (~ 2.5 min of data). The right-most panel shows the t-value distribution for NORDIC, Standard reconstruction, and Standard reconstruction with 2 mm Full Width Half Maximum (FWHM) spatial smoothing, within a region of interest (ROI) created using the largest cluster from 8 runs of the data with Standard reconstruction and threshold of t=3.297, for the target *vs.* surround contrast. This ROI contained 2762 voxels. This figure demonstrates the benefit of NORDIC denoising with supra-millimeter acquisition protocols at 3 Tesla. The HCP acquisition protocol implemented here represents a gold standard for 3T fMRI; therefore, these results further highlight the wide applicability and relevance of NORDIC denoising.

employed in imaging in general, and in fMRI in particular, to improve the image signal-to-noise ratio (SNR) albeit at the expense of degrading spatial resolution; the impact of this filter on the t-statistics is similar to NORDIC, though NORDIC is somewhat better than this specific amount of smoothing. However, unlike spatial smoothing, NORDIC achieves this improvement without degradation of spatial resolution (data presented in the main manuscript; also see discussion about Supplementary Fig. 8).

Note that the functional maps (t-statistics maps with thresholding) obtained with NORDIC using a single run is virtually identical on a pixel-by-pixel basis to that obtained with Standard 4 runs.

Looking into the future, one of the major impacts of the NORDIC approach will likely be to enable 3 Tesla studies at higher spatial resolutions relative to what has been achievable to date. We demonstrate this potential in Supplementary Fig. 6 below, using 3 Tesla data acquired with 1.2



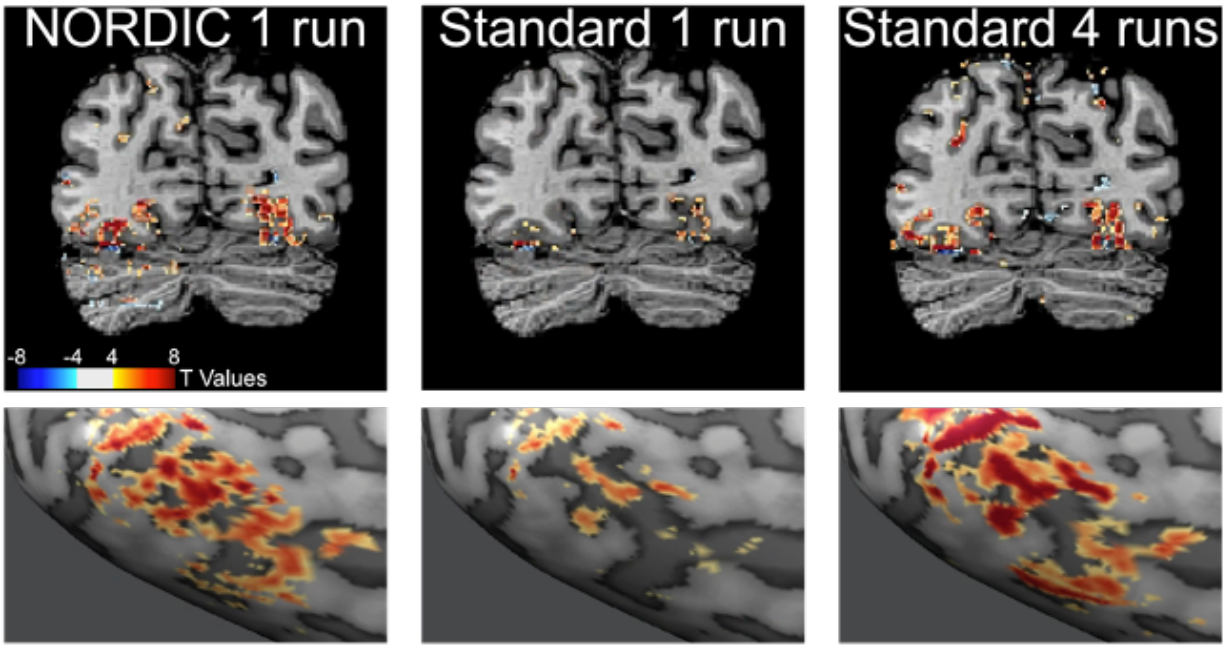
Supplementary Figure 6: The effect of NORDIC denoising on an fMRI study carried out at 3 Tesla using 1.2 mm isotropic resolution (TR 2100 ms; MB factor= 4; iPAT 2). The figure show t-maps with a t-threshold of ≥ |4.5| for the contrast target (red) > surround (blue). Because the higher resolution lengthens the echotrain length with deleterious contributions to image distortions and signal drop out, this acquisition employed parallel imaging along the phase encode direction (iPAT=2), and reduced the parallel imaging along the slice direction by using MB=4 as opposed MB=8 of the standard HCP protocol (Supplementary Fig.1). The stimulus and the presentation paradigm were the same as in Supplementary Fig.1 and identical to that implemented in the main study (see Methods), where flickering gratings for target or surround were presented for 12 seconds blocks interleaved by 12 seconds fixation periods. The figure shows t-maps for the contrast target (red) > surround (blue) in volume space, for a representative slice in the visual cortex (top row) and in inflated cortical space (bottom row). Approximately 5 Standard runs (~12 minutes of data) are required to achieve the extent of activation comparable to a single NORDIC run (~2.5 minutes of data). Given the significant differences in acquisition parameters relative to the HCP 3T data (Supplementary Fig.1), these results demonstrate the benefits of NORDIC in highly different MR context, further advocating for the generalizability of the technique.

mm isotropic resolution, again using the same visual stimulation paradigm employed in the manuscript and for the data in Supplementary Fig. 5.

The challenge of this higher resolution at 3 Tesla becomes evident with the functional map obtained from a single run processed with the Standard approach (Supplementary Fig. 6, middle column); this functional map is missing most of the activated territory evident in the data shown in Supplementary Fig. 5. However, the large activated territories seen in Supplementary Fig. 5 are largely recovered in NORDIC reconstruction of a single run (Supplementary Fig. 6, left column) or with Standard reconstruction after concatenating five runs, representing a much longer (~12 min) data acquisition (Supplementary Fig.6, right most column).

ii) EVENT RELATED Designs with COVENTIONAL (Supra-millimeter) and HIGH (sub-millimeter) SPATIAL RESOLUTIONS and using complex COGNITIVE tasks (Face detection and gender discrimination):

In order to demonstrate the gains achieved by NORDIC are applicable at 7T data with i) more standard (i.e. $\gtrsim$ 1 mm isotropic) spatial resolutions, ii) different stimuli or tasks, iii) different

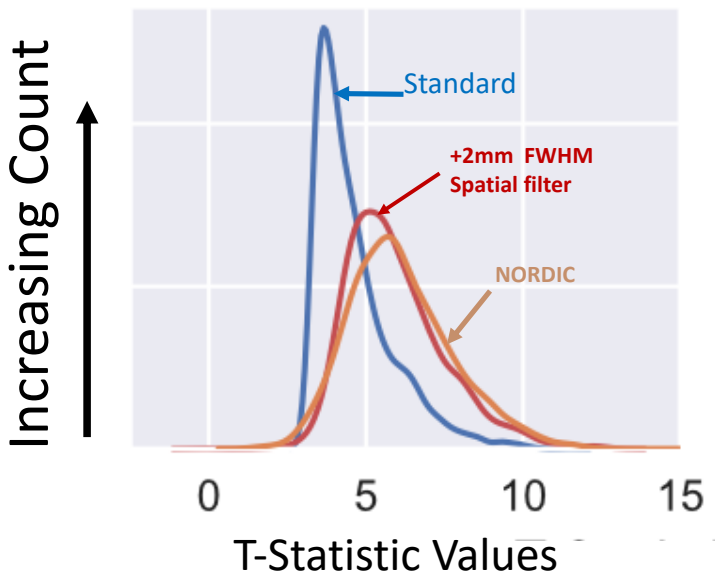


Supplementary Figure 7: The impact of NORDIC denoising on 7T data acquired with standard HCP 7T fMRI protocol (i.e. 1.6 mm iso voxels; TR 1000 ms; MB 5; iPAT 2). Here, to demonstrate the benefits of NORDIC denoising across different experimental contexts, we used 2 seconds stimulus-on, 2 seconds stimulus-off, fast event related design. A 12 s stimulus-off baseline period was also utilized at the beginning and at the end of the rapid, 2 second on-off alternating epochs. During the stimulus on periods, the participants viewed degraded images of faces (i.e. ranging from 0% to 40% image phase coherence in steps of 10% increments) while preforming a face detection task. The figure shows t-maps ($t \geq |4|$) for the mean activation of all conditions vs. the mean of all baseline periods when nothing was shown; the top row shows this activation map in volume space for a single slice and the lower row displays it on the inflated cortical surface in the fusiform face area. Four Standard runs (~13 minutes of data) are required to achieve the extent of activation comparable to a single NORDIC run (~3 minutes and 20 seconds of data). The HCP acquisition protocol implemented here represents a gold standard for supra-millimeter 7T fMRI; therefore, these results highlight the wide-ranging advantages conferred by NORDIC denoising across acquisition and experimental protocols.

paradigms, and even iv) different regions of the cortex, we turned again to the HCP as a "gold standard". The HCP had a 7 Tesla component[1,2], and this component employed 1.6 mm isotropic resolution, TR=1 s, SMS/MB acquisition with MB=5, iPAT (undersampling in phase encoded direction) = 2. We used these acquisition parameters at 7T with a face detection paradigm where participants viewed degraded images of faces (ranging from 0% to 40% image phase coherence in steps of 10% increments) while preforming a face detection task; the visual stimulation paradigm employed was a *fast event related* design where every image was presented for 2 seconds, followed by a 2 second fixation period (with 10% blank trials). The activation maps portrayed in Supplementary Fig. 7 show the T value maps with a t-threshold of $t \geq |4|$ for the mean activation to all 5 visual conditions. In this case the cortical region shown is the face fusiform area in the inferior temporal lobe.

Again, we see that single run processed with NORDIC provides a functional map that is virtually equivalent to the map obtained using multiple runs (in this case 4) concatenated and processed with Standard reconstruction algorithms, and both are far superior to what is seen with a single run processed with Standard reconstruction (Supplementary Fig.7).

Supplementary Figure. 8 compares the effect of NORDIC on the t-statistics against the effects of spatial filtering (smoothing) for an event related paradigm similar to that shown in Supplementary Fig.7, but acquired with higher spatial resolution (0.8 mm isotropic) using face presentations during a gender discrimination task. Supplementary Fig.8 illustrates that applying a 0.8 mm



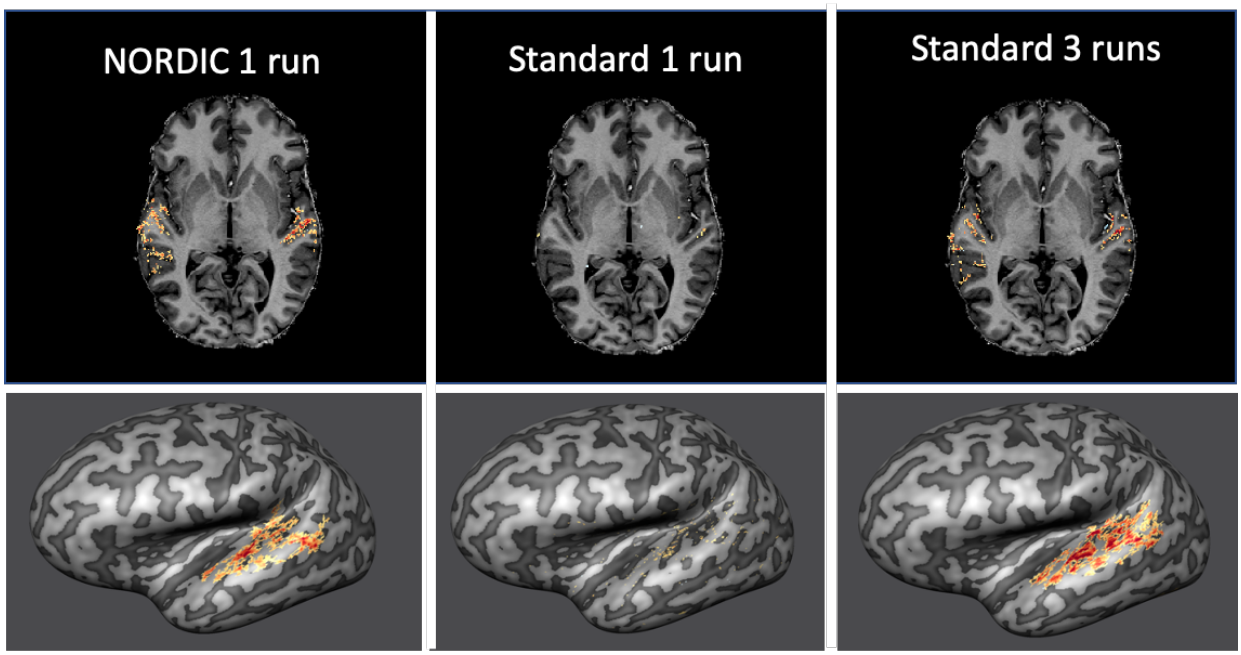
Supplementary Figure 8: T-values extracted from an event related task design. Data were collected with an anterior to posterior phase encoding direction (7 Tesla, TR 1.4s, Multiband 2, iPAT 3, 6/8ths Partial Fourier, 0.8mm isotropic, TE 27.4ms, Flip Angle 78°, Bandwidth 1190Hz) and covered the occipital pole and ventral temporal cortex. Six runs of data were collected. Stimuli consisted of faces *versus* fully phase scrambled variants of the same faces. Stimuli were presented for 2 seconds, with a 2 second interstimulus interval. T-statistics for various methods were extracted using an ROI produced from 6 runs of the Standard reconstruction (blue) contrasting faces vs. scrambled. NORDIC (orange) produces T-statistic values that exceed the Standard dataset (blue). NORDIC t-statistics is more comparable to, but slightly better than that observed when the original data is spatially smoothed with a 0.8mm FWHM gaussian filter (red); however, NORDIC achieves the improvement in t-statistics without spatially blurring the image.

FWFM spatial smoothing on the raw data improves the t-statistics significantly, as in the case of the 3T data shown in Supplementary Fig. 5, albeit with the effective spatial resolution degraded by approximately a factor of two in this case. NORDIC, on the other hand, performs comparable to or slightly better than spatial filtering with respect to improvements in the t-statistics (Supplementary Fig. 8), but does not degrade spatial resolution (see data presented in the main body, and also Supplementary Fig. 10 and related discussion). Spatial filtering improves image SNR by averaging signals over regions that are larger than those determined by the voxel dimensions set by the acquisition parameters, therefore greatly compromising the spatial specificity of fMRI.

NORDIC instead can be thought of as having the same positive effect on the ability to detect stimulus/task induced signal changes in fMRI as spatial filtering *without* the deleterious consequences of image blurring that come with spatial filtering of images.

iii) fMRI with AUDITORY STIMULUS

As a demonstration of yet another different sensory stimulus and different cortical region, we present data obtained with auditory stimuli. The results again are comparable to those presented in the main text as well as the additional data included in this supplementary material as shown in Supplementary Figs. 5 through 8. Namely, multiple runs obtained over significantly longer data acquisition times are needed with Standard reconstruction to achieve the extent of activation comparable to a single NORDIC run.

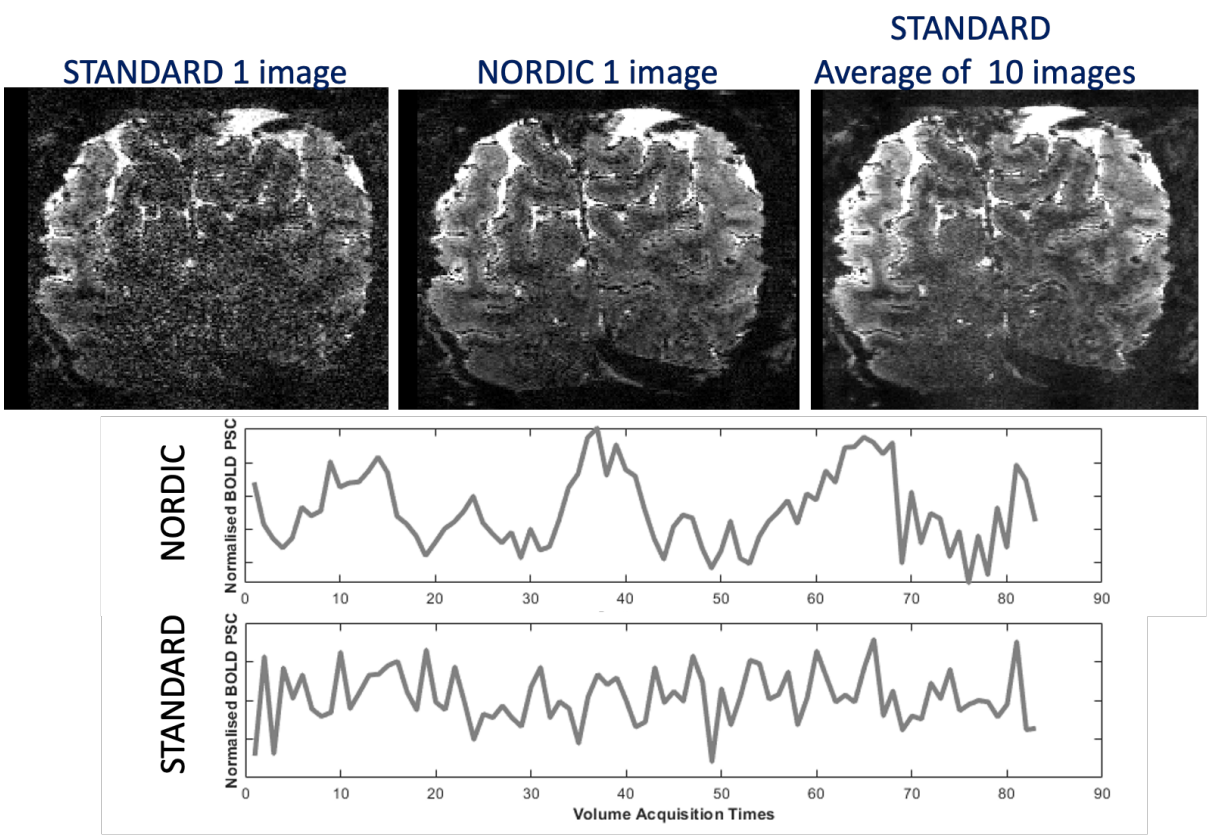


Supplementary Figure 9: The impact of NORDIC denoising on 7T data acquired in response to sound presentation (0.8 mm iso voxels; TR 1600 ms; MB 2; iPAT 3; 42 slices). Sounds (simple tones) were presented following a slow event related design (each sound lasting 1 second were presented with an average inter trial interval of 10 seconds) while participants passively listened to them. The figure shows t-maps for the main effect of sound presentation (i.e. sound vs. silence; thresholded at $t > 2.7$) in volume space (top row – one representative transversal slice) and in inflated cortical space (bottom row – left hemisphere). Note, 3 Standard runs (approximately 18 minutes of data) are required to achieve the extent of activation comparable to a single NORDIC run (approximately 6 minutes of data). These results highlight advantages conferred by NORDIC denoising beyond application to visual experiments.

iv) GOING to EVEN HIGHER RESOLUTIONS

The spatial resolution of an MR image is ultimately limited by SNR. Since the signal magnitude of an MR voxel is proportional to the voxel volume, the relevant parameter to consider for discussions of SNR *vs.* spatial resolution is the voxel volume; this allows the considerations to be generalized to include voxels with non-isotropic dimensions. The 0.8 mm isotropic resolution employed in the data presented in the manuscript and also in Supplementary Figs. 8 and 9 corresponds to a voxel volume of ~0.5 μL.

The 0.8 mm isotropic resolution is typical of spatial resolutions employed in the ever growing 7T efforts to investigate brain function down to the scale of mesoscopic organizations. Although such studies have led to successful functional mapping of cortical columns and layers, and other fine-scale organizations[3-5], they are highly SNR starved and hence a good target for NORDIC applications. In reality, however, even 0.8 mm isotropic resolution is still marginal if not outright coarse for accurately mapping mesoscopic scale organizations. This limitation is reflected in the first report of the BRAIN Initiative Working group to the Director of NIH [6,7], which calls for



Supplementary Figure 10: Top row shows a single slice from a single time point in the consecutively acquired volumes forming an fMRI time series. The data were acquired using a 3D gradient echo (GE) EPI approach with a spatial resolution of 0.5 mm isotropic voxels (3D GE EPI; 40 slices; iPAT 3; TR 83 ms; Volume acqusition time (VAT) 3652 ms). The total data acqusition time for the fMRI time series was ~ 5 minutes (i.e. 2 visual conditions - center and surround gratings – each presented 3 times for 24 seconds, and each followed by a 24 seconds fixation period. The top row shows the image of selected single slice from the Standard reconstruction (left), the same single slice from NORDIC reconstructed data (middle), and the average of 10 images of the same slice for the Standard reconstruction (right). The bottom 2 panels show the timecourse of a single medial voxel for NORDIC (top) and Standard (bottom) reconstructions.

transformative technical developments in non-invasive MRI based brain mapping methods to measure the "*functional activity of the human brain at multiple scales—neuronal ensembles, circuits, and larger scale networks ('circuits of circuits')*" and challenges the MR community to achieve whole brain imaging studies with of 0.1 μL or smaller voxel volumes (see p 71 and 75, BRAIN2025 [6]).

Supplementary Fig.10 shows 0.5 mm isotropic fMRI data (0.125 μL voxel volume) obtained using ~5 min of data collection using a visual stimulation paradigm of flickering gratings comparable to that employed in the main body and in Supplementary Figs 5 and 6. Because high resolution data takes longer to acquire, the visual stimulation and fixation periods were increased to 24 seconds each (from the 12 seconds employed previously), so as to allow sufficient fMRI time points to be collected during these periods. This doubled the duration of a single run from ~2.5 min to ~ 5min.

Supplementary Fig.10, top row, shows a single coronal slice in the visual cortex from one of the repetitively acquired volumes in the fMRI time series. Processed with the Standard reconstruction, the image of this slice (Supplementary Fig. 10, left column) demonstrates that this resolution yields extremely noisy images, which are unusable for functional mapping. However, the single image after NORDIC reconstruction (Supplementary Fig. 10, middle column), and average of 10 images from the Standard reconstruction (Supplementary Fig. 10, right column) look virtually identical, even at the level of each small black dots which arise from small venous blood vessels traversing the slice approximately perpendicularly to the slice orientation. In addition to the smoothing metrics given in the main text, these very high resolution data also demonstrate clearly that NORDIC does not induce meaningful smoothing, as every detail visible in the Standard Average of 10 images, such as gray-white matter boundaries, their tortuosity, the venous vessels (visible as dark lines and dots) etc. are also clearly and identically seen in the NORDIC single image. Also shown in Supplementary Fig. 10, is a time course from a single voxel in the visual cortex. The stimulus induced signal changes in a single medial voxel are simply not detected in the Standard reconstruction at this very low SNR, 0.5 mm isotropic resolution data, but becomes obviously visible with the use of NORDIC.

Supplemental References

1. Ugurbil, K., Xu, J., Auerbach, E. J., Moeller, S., Vu, A. T., Duarte-Carvajalino, J. M., Lenglet, C., Wu, X., Schmitter, S., Van de Moortele, P. F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D. A., Smith, S. M., Miller, K. L., Sotiropoulos, S. N., Jbabdi, S., Andersson, J. L., Behrens, T. E., Glasser, M. F., Van Essen, D. C., Yacoub, E. & Consortium, W. U.-M. H. *Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project*. (2013) Neuroimage **80**, 80-104.
2. Glasser, M. F., Smith, S. M., Marcus, D. S., Andersson, J. L., Auerbach, E. J., Behrens, T. E., Coalson, T. S., Harms, M. P., Jenkinson, M., Moeller, S., Robinson, E. C., Sotiropoulos, S. N., Xu, J., Yacoub, E., Ugurbil, K. & Van Essen, D. C. *The Human Connectome Project's neuroimaging approach*. (2016) Nat Neurosci **19**, 1175-1187.
3. Ugurbil, K. *Imaging at ultrahigh magnetic fields: History, challenges, and solutions*. (2018) Neuroimage **168**, 7-32.
4. De Martino, F., Yacoub, E., Kemper, V., Moerel, M., Uludag, K., De Weerd, P., Ugurbil, K., Goebel, R. & Formisano, E. *The impact of ultra-high field MRI on cognitive and computational neuroimaging*. (2018) Neuroimage **168**, 366-382.
5. Dumoulin, S. O., Fracasso, A., van der Zwaag, W., Siero, J. C. W. & Petridou, N. *Ultra-high field MRI: Advancing systems neuroscience towards mesoscopic human brain function*. (2018) Neuroimage **168**, 345-357.
6. BRAIN Initiative Working Group. *BRAIN 2025: A Scientific Vision*, <http://braininitiative.nih.gov/pdf/BRAIN2025_508C.pdf> (2014).
7. Jorgenson, L. A., Newsome, W. T., Anderson, D. J., Bargmann, C. I., Brown, E. N., Deisseroth, K., Donoghue, J. P., Hudson, K. L., Ling, G. S., MacLeish, P. R., Marder, E., Normann, R. A., Sanes, J. R., Schnitzer, M. J., Sejnowski, T. J., Tank, D. W., Tsien, R. Y., Ugurbil, K. & Wingfield, J. C. *The BRAIN Initiative: developing technology to catalyse neuroscience discovery*. (2015) Philos Trans R Soc Lond B Biol Sci **370**.