ARTICLE

Check for updates

Lowering the thermal noise barrier in functional brain mapping with magnetic resonance imaging

Luca Vizioli [1,2,5 ✉], Steen Moeller[1,5], Logan Dowdle [1,2], Mehmet Akçakaya [1,3], Federico De Martino [1,4], Essa Yacoub[1] & Kamil Uğurbil [1 ✉]

Functional magnetic resonance imaging (fMRI) has become an indispensable tool for investigating the human brain. However, the inherently poor signal-to-noise-ratio (SNR) of the fMRI measurement represents a major barrier to expanding its spatiotemporal scale as well as its utility and ultimate impact. Here we introduce a denoising technique that selectively suppresses the thermal noise contribution to the fMRI experiment. Using 7-Tesla, high-resolution human brain data, we demonstrate improvements in key metrics of functional mapping (temporal-SNR, the detection and reproducibility of stimulus-induced signal changes, and accuracy of functional maps) while leaving the amplitude of the stimulus-induced signal changes, spatial precision, and functional point-spread-function unaltered. We demonstrate that the method enables the acquisition of ultrahigh resolution (0.5 mm isotropic) functional maps but is also equally beneficial for a large variety of fMRI applications, including supra-millimeter resolution 3- and 7-Tesla data obtained over different cortical regions with different stimulation/task paradigms and acquisition strategies.

[1] Center for Magnetic Resonance Research (CMRR), University of Minnesota, Minneapolis, MN, USA. [2] Department of Neurosurgery, University of Minnesota, Minneapolis, MN, USA. [3] Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA. [4] Faculty of Psychology and Neuroscience, Department of Cognitive Neurosciences, Maastricht University, Maastricht, the Netherlands. [5]These authors contributed equally: Luca Vizioli, Steen Moeller. ✉email: luca.vizioli1@gmail.com; ugurb001@umn.edu

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 the ultrahigh magnetic field (UHF) of 7 Tesla (7 T)[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[7–9]. 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]. Such studies provide unique opportunities for investigating the organizing principles of the human cortex at the mesoscopic scale, thus bridging the gap between invasive electrophysiology and optical imaging studies and non-invasive human neuroimaging.

Despite these successes, however, the signal-to-noise and the functional contrast-to-noise ratios (SNR and fCNR, respectively) of fMRI measurements remain relatively low. This represents a major impediment to expanding the spatiotemporal scale of fMRI applications as well as the utility, interpretation, and ultimate impact of fMRI data.

What is considered "noise" in an fMRI time series is a complex question. Thermal noise associated with the MR detection[10,11], arising either from the electronics and/or the sample, is an important 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[12]. 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[13,14]. It is the *dominant* contribution at ~0.5 µL voxel volumes (e.g. ~0.8 mm isotropic dimensions) typically employed in high resolution 7 T fMRI studies; it remains dominant at 7 T up to ~10 µL voxel volumes, gradually plateauing beyond that[13,14]. However, even with 3 mm isotropic resolution (i.e. 27 µL voxel volume) and relatively long TR acquisitions, thermal noise was estimated to be a significant contributor to fMRI time series at 7T[15]. At lower magnetic fields like 3 T, where this type of noise becomes more conspicuous, and where typical fMRI resolutions employed are ≲3 mm, it would be a substantial contributor in virtually all fMRI studies[13,14].

In this paper, we tackle these SNR and fCNR limitations using a denoising technique—namely, NOise Reduction with DIstribution Corrected (NORDIC) PCA. NORDIC operates on repetitively acquired MRI data and only removes components that cannot be distinguished from zero-mean Gaussian distributed noise; as such, the method targets the suppression of thermal noise and not the structured, non-white noise caused by respiration, cardiac pulsation, and spontaneous neuronal activity (e.g. refs. [16–19] and references therein).

High-resolution 7 Tesla data, as well as data obtained with more conventional, supra-millimeter resolution at 3 and 7 T using several different task/stimulus and acquisition strategies, demonstrate that major gains are achievable under a wide variety of experimental conditions with NORDIC in gradient-echo (GE) BOLD fMRI without introducing image blurring. Based on these findings, the approach is expected to markedly widen the scope and applications of fMRI in general, and high spatial and/or temporal resolution fMRI in particular.

Results

The fMRI data, acquired with GE simultaneous multi-slice (SMS)/multiband (MB) echo planar imaging (EPI)[20,21], were reconstructed either by the algorithms provided with the MR scanner (referred throughout this work as "Standard") or by the NORDIC PCA method (see the "Methods" section for a detailed description) using the raw k-space files produced by the scanner (referred to as "NORDIC").

The bulk of the analyses were performed on data acquired on four subjects with a variant of a widely used, 0.8 mm isotropic resolution 7 T protocol (see the "Methods" section) based on a block design visual stimulation paradigm (Fig. 1A); these analyses are presented in this section. However, to ensure the generalizability of our results and the versatility of the NORDIC approach, we present as Supplementary Material evaluations of NORDIC on fMRI across acquisition parameters, field strengths (i.e. 3 and 7 T), cortical regions, and stimulation paradigms, bringing the total number of datasets to $N = 10$. All data sets showed converging results.

We used a block design, visual stimulation paradigm comparable to that implemented in Shmuel et al.[22] with minor modifications: It consisted of retinotopically organized target and surround stimuli presented in alternating stimulus-on and stimulus-off epochs (Fig. 1A). Each "run" consisted of six stimulus-on epochs, three each for target and surround stimuli. We acquired 8 experimental runs in 6 subjects (4 at 7 T and 2 at 3 T, the latter presented as Supplementary Material); 2 of these runs were used to identify the retinotopic representation of the target in V1, computed by contrasting the target versus 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 the "Methods" section).

NORDIC vs. Standard MR images. Fig. 1B illustrates an example slice for Standard and NORDIC reconstructed GE-EPI images for two subjects before any preprocessing for fMRI analysis was applied. An improvement is visually perceptible for NORDIC images, especially in the central regions where the g-factor noise amplification would be particularly elevated. Subtraction of Standard from NORDIC processed image of a single slice from a single time point in the fMRI times series displayed only noise without any features of the image or edge effects; when such a difference was calculated for all time points in the fMRI time series and averaged, the result was equivalent to the g-factor map (Supplementary Fig. 1). These observations are consistent with NORDIC suppressing only random noise without impacting the image.

Figure 1C shows temporal SNR (tSNR) maps averaged across all eight runs for two exemplar subjects and slices. The average tSNR across *all* the voxels in the brain was more than 2-fold larger for NORDIC (S1$_{tSNR}$: 27.34 ± 2.26 (std); S2$_{tSNR}$: 33.01 ± 2.26 (std); S3$_{tSNR}$: 43.31 ± 1.52 (std); S4$_{tSNR}$: 26.61 ± 2.32 (std)) compared to Standard (S1$_{tSNR}$: 13.28 ± 0.19 (std); S2$_{tSNR}$: 13.97 ± 0.05 (std); S3$_{tSNR}$: 16.1 ± 0.26 (std); S4$_{tSNR}$: 14.12 ± 0.23 (std)) images. Paired sample t-tests carried out across all 8 runs,
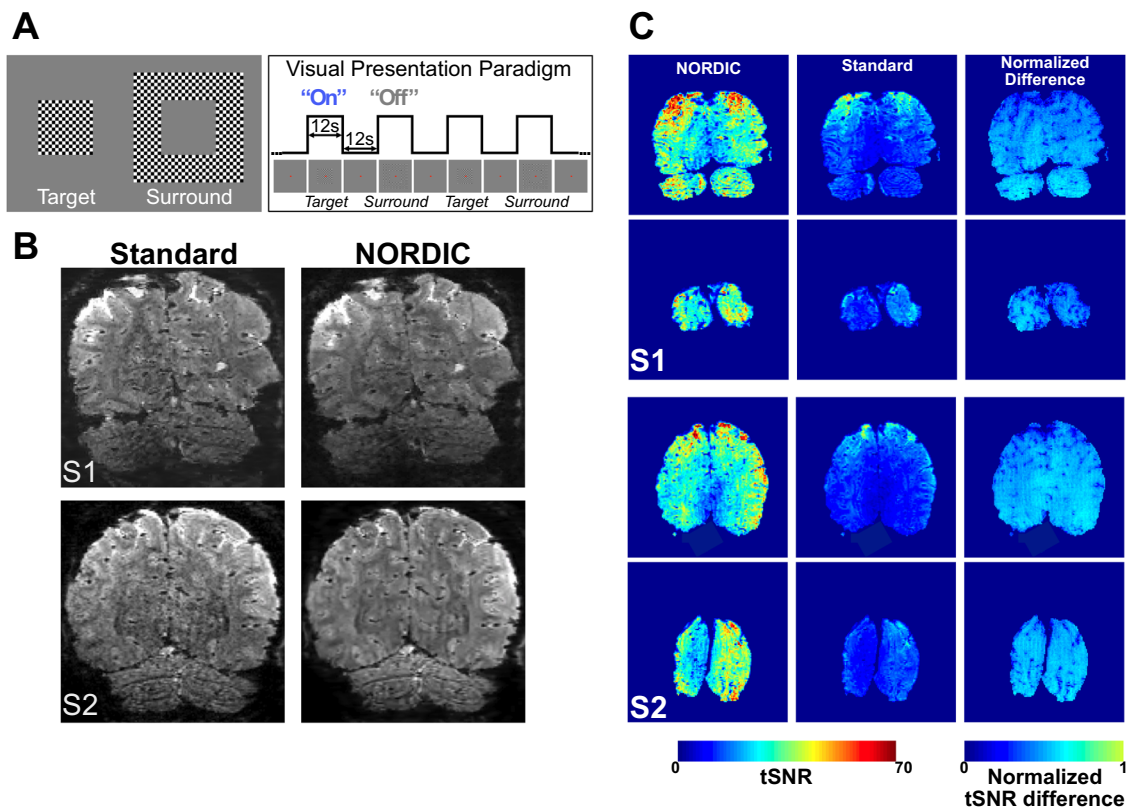
Fig. 1 Stimuli and paradigm, epi images, and tSNR. Panel **A** depicts the visual stimuli (left) used and a schematic of the visual presentation paradigm (right). Panel **B** shows an example slice from a single volume extracted from an fMRI time series for Standard (left column) and NORDIC (right column) reconstructions before any preprocessing, for two subjects S1 and S2. Panel **C** shows average (across all 8 runs) brain temporal signal-to-noise ratio (tSNR) maps of 2 exemplar slices in 2 representative subjects (S1 and S2) for NORDIC (left) and Standard (center) reconstructions and the normalized difference between the 2. The last was computed by performing (tSNR$_{NORDIC}$ -tSNR$_{STANDARD}$)/tSNR$_{NORDIC}$). The slices chosen represent one of the anterior-most slices in the covered volume, and an occipital slice that includes a portion of the target ROI in V1.

independently per subject, indicated that for all subjects, the average tSNR for NORDIC was significantly larger ($p < 0.01e^{-5}$) than that for Standard images. Improvements in tSNR with NORDIC in individual runs are shown in Supplementary Figs. 2 and 3.

Functional images. The impact of NORDIC on functional maps was evaluated by comparing a single run processed with NOR-DIC against the concatenation of multiple runs of the Standard reconstruction (see the "Methods" section). Figure 2 illustrates functional maps on the inflated surface of one hemisphere, contrasting the target versus the surround condition thresholded at $|t| \geq 5.7$ for four subjects. For two subjects, representative single-run functional maps are also shown for two different t-thresholds and on the anatomical image of a slice in Supplementary Fig. 4. At the same t-threshold, the extent of activation achievable with a single NORDIC run was comparable or better than that obtained by concatenating 3–5 Standard runs.

Similar results are presented in Supplementary Material for supra-millimeter 3 and 7 T fMRI data obtained with visual stimulation and face recognition paradigms (Supplementary Figs. 8–11), and for 0.8 mm 7 T data obtained with auditory stimulation (Supplementary Fig. 12); two of these datasets (Supplementary Figs. 10 and 11) were acquired with an event-related paradigm.

Consistent with the data displayed in Fig. 2, the t-values examined further in two subjects (S1 and S2) were significantly larger for NORDIC ($p < 0.05$) than its Standard counterpart (Fig. 3A, E) within the target ROI, as determined with linear mixed models carried out independently per subject. When the t-value distribution for the target > 0 contrast was analyzed for three ROIs (Supplementary Fig. 6), it was found to be shifted to higher values in each individual run for the target ROI; for the two other ROIs in regions where stimulus-evoked responses should not exist, it was essentially unaltered, demonstrating that NORDIC does not perturb t-values where it should not.

Percent signal change (PSC) within the target ROI as the mean of all voxels and at the single-voxel level are presented in Fig. 3B, F, respectively. The stimulus-induced PSC was highly comparable across reconstruction types; linear mixed models carried out independently per subject (with the individual runs as a random effect) showed no significant ($p > 0.05$, Bonferroni corrected) differences in PSC amplitudes across reconstructions for all runs.

Figure 3C depicts the standard deviation (20% trimmed mean across voxels within target ROI[23]) 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, indicating greater stability of NORDIC PSC single-trial estimates among the different stimulus epochs within a run.

The equivalency of PSC amplitudes for NORDIC and Standard reconstructions are further illustrated using images in Fig. 4 and Supplementary Fig. 5. In addition, hold-out data analysis was carried out with PSC estimates (Fig. 4); 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
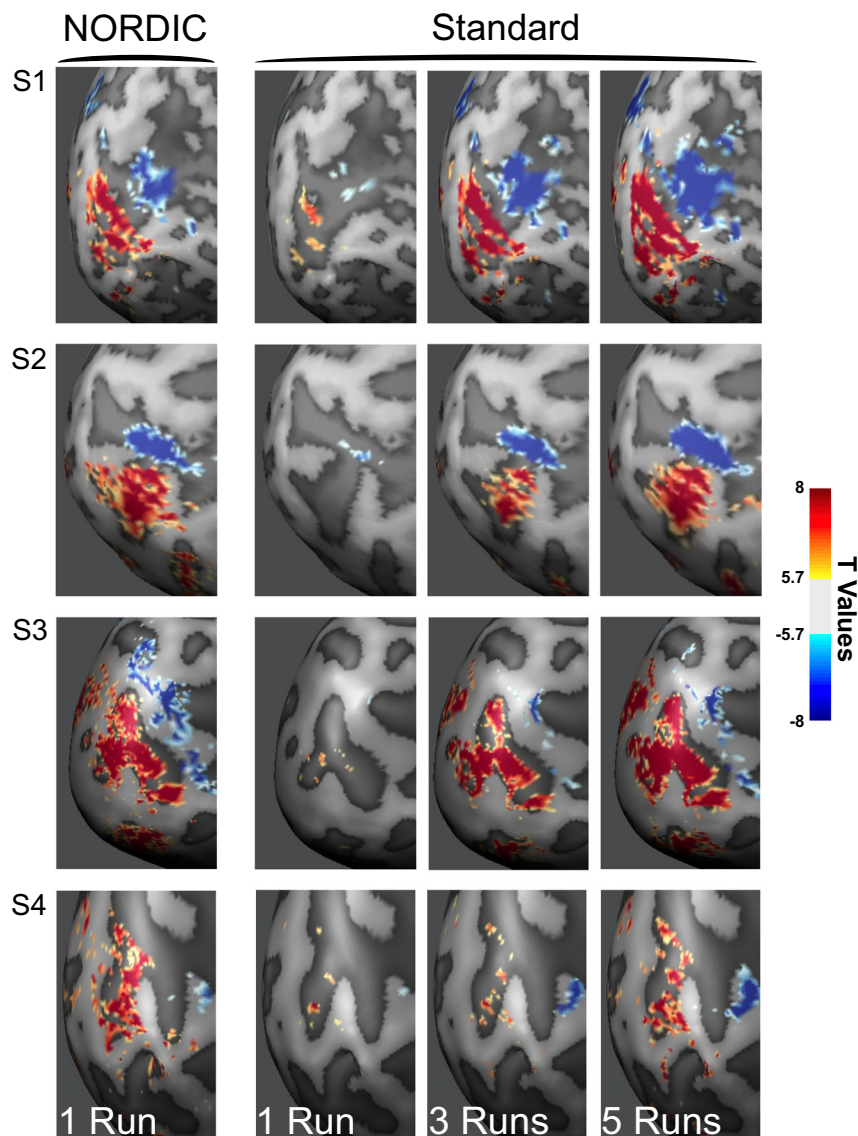
Fig. 2 NORDIC vs. Standard *t*-Maps. Leftmost panel shows functional images as *t*-maps (target > surround) thresholded at |*t*| ≥ 5.7 for a single NORDIC processed run, and for 1, 3, and 5 Standard processed runs combined, for subject 1 (S1), subject 2 (S2), subject 3 (S3) and subject 4 (S4).

at a single voxel level. The precision of PSC estimates, computed as cross-validated R2 for single run GLMs, was higher (Fig. 4A, third row) for the NORDIC compared to the Standard reconstruction. Paired sample 2-sided *t*-test carried out across cross-validations folds showed that within the target ROI average $R2$ (see the "Methods" section) was significantly ($p < 0.01$ Bonferroni corrected) higher for NORDIC (S1: NORDIC mean $R2 = 36.43$ (ste $= 9.81$); Standard mean $R2 = 22.2$ (ste $= 5$); S2: NORDIC mean $R2 = 25.36$ (ste $= 5.74$); Standard: mean $R2 = 10.62$ (ste $= 3.2$) and Fig. 4B, bar graphs), indicating again higher precision of PSC estimates and their stronger predictive value for NORDIC.

Figure 5 shows the functional point spread (PSF) measurements on the cortical surface calculated following previous work[22] using NORDIC and Standard images (see the "Methods" section) from two subjects: briefly, the approach defines the boundary between the target and the surrounding stimuli as those voxels showing a *differential* functional response close to 0 (Fig. 5A, 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[22] (Fig. 5B). The full-width at half maximum (FWHM) of the Gaussian represents the functional PSF[22]. With NORDIC, the average PSFs (across traces) were 1.04 mm (std: 0.19) and 1.22 mm (std: 0.51), for subjects 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.

In addition, we estimated the global smoothness of individual GE-SMS/MB-EPI images in the fMRI time series using AFNI (3dFWHMx function)[24], 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. The FWHM from this mixed model estimate, averaged over four subjects, before and after data preprocessing (see the "Methods" section) for the Standard reconstruction was 0.92 ± 0.002 and 0.94 ± 0.05 mm, respectively; for the NORDIC reconstruction, these values were 0.93 ± 0.02 and 0.94 ± 0.05 mm
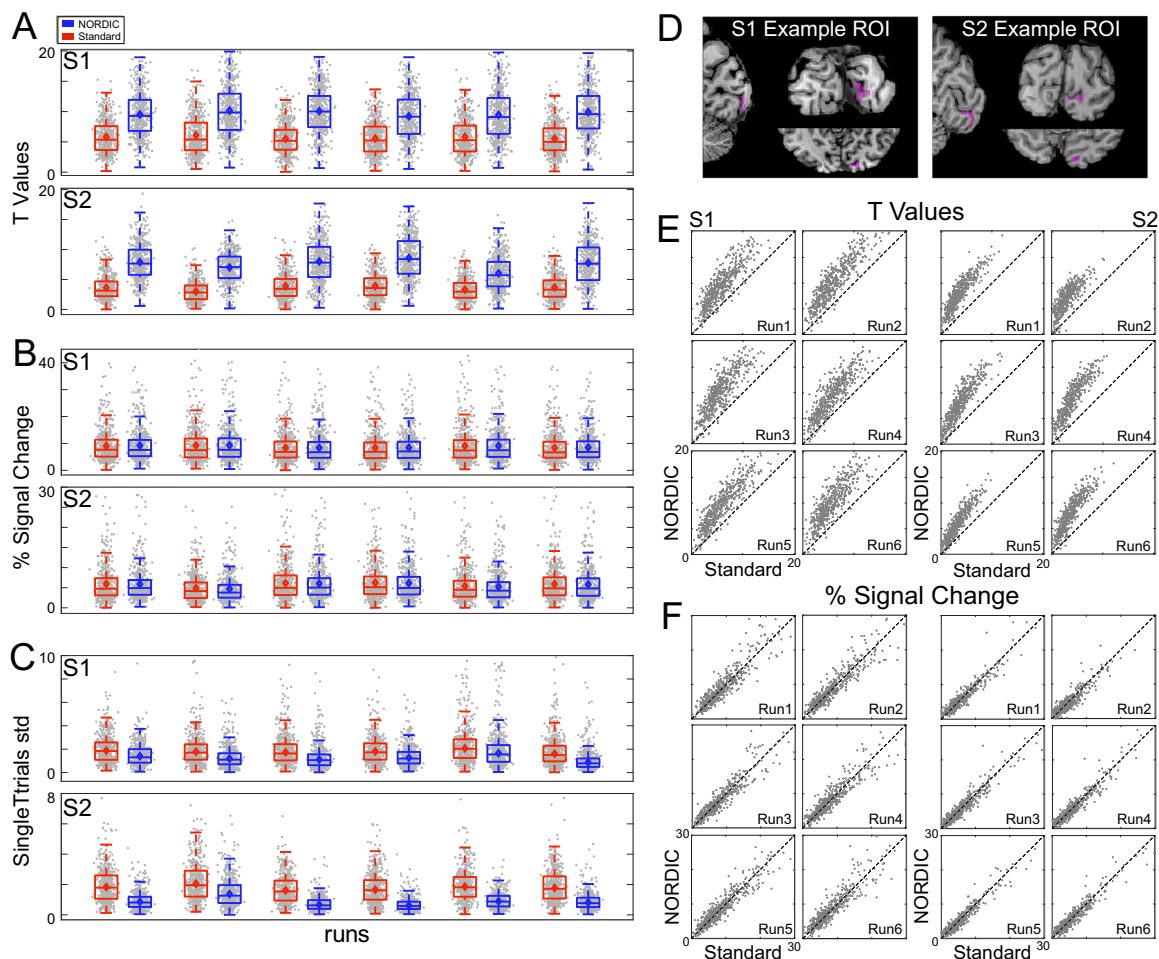
Fig. 3 Voxel responses within target ROI. Panel **A** shows the single-run (arranged over the x-axis) t-values (activity elicited by the target >0) induced by the target stimulus for Standard (red) and NORDIC (blue) data. Panel **B** is the same as panel **A**, but for beta weights (transformed into percent signal change). Panel **C** shows the single-run standard deviation computed across single-trial PSC beta estimates elicited by the target condition. For these three panels, gray dots represent responses to single voxels with the target ROI (497 for S1 and 461 for S2). The box-and-whisker plots, computed across all ROI voxels, represent the interquartile range (IQR—with box limits being the upper and lower quartiles), with the whiskers extending 1.5 times the IQR or to the largest value. The horizontal lines within the boxplot represent the median, while the diamond the mean across voxels. Panel **D** shows the target ROI, representing the left retinotopic representation of the target in V1 for 2 exemplar subjects in all three planes. Panel **E** shows the single runs, single-voxel scatterplots for t-values (activity elicited by the target > 0), for Standard (x-axis), and NORDIC (y-axis). Panel **F**: same as panel **E** for the beta percent signal change responses to the target condition. Source data are provided as source Data files.

(Fig. 5C). A linear-mixed model carried out across subjects and runs indicated nonmeaningful differences in smoothness estimate ($p > 0.05$) between Standard and NORDIC.

Figure 6 shows 0.5 mm isotropic resolution fMRI data (0.125 μL voxel volume) obtained using the target/surround visual stimulation paradigm (see the "Methods" section and also Supplementary Fig. 14). Figure 6A and Supplementary Fig. 14A display 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 is very noisy and practically unusable for functional mapping. However, the single image after NORDIC reconstruction and average of 10 images from the Standard reconstruction look virtually identical; these very high-resolution data also demonstrate clearly that NORDIC does not induce smoothing (see expanded panels in Supplementary Fig. 14 and also Panel D in Supplementary Fig. 16).

Functional maps from the 0.5 mm data computed for the target > surround contrast using the 8 concatenated runs (i.e. ~44 minutes of data for the two stimulus conditions and interleaving baseline periods) are shown superimposed on T_1-weighted anatomical images (Fig. 6B) and the flattened cortex (Fig. 6C). These functional data do not have any spatial smoothing or masking applied to them. Little activation is detected with Standard reconstruction. Localized and highly precise BOLD activation, allowing differentitation of adjacent sulcus banks (Fig. 6B) are observed for NORDIC images. Consistent with these observations, stimulus-evoked signal changes in the fMRI time course at a single voxel level were virtually undetectable in Standard reconstruction but obviously visible with the use of NORDIC (Supplementary Fig. 14B).

The NORDIC method should be equally applicable for resting state fMRI (rsfMRI) that is used extensively to evaluate functional connectivity. We present a preliminary analysis on one subject at 3 T confirming this expectation (Supplementary Fig. 13).

As previously mentioned, the "Standard" reconstruction employed in these comparisons is the one provided by the vendor of the scanner. For NORDIC, prior to denoising, the same experimentally acquired k-space data was exported and had to be processed "offline" for EPI and GRAPPA reconstructions using
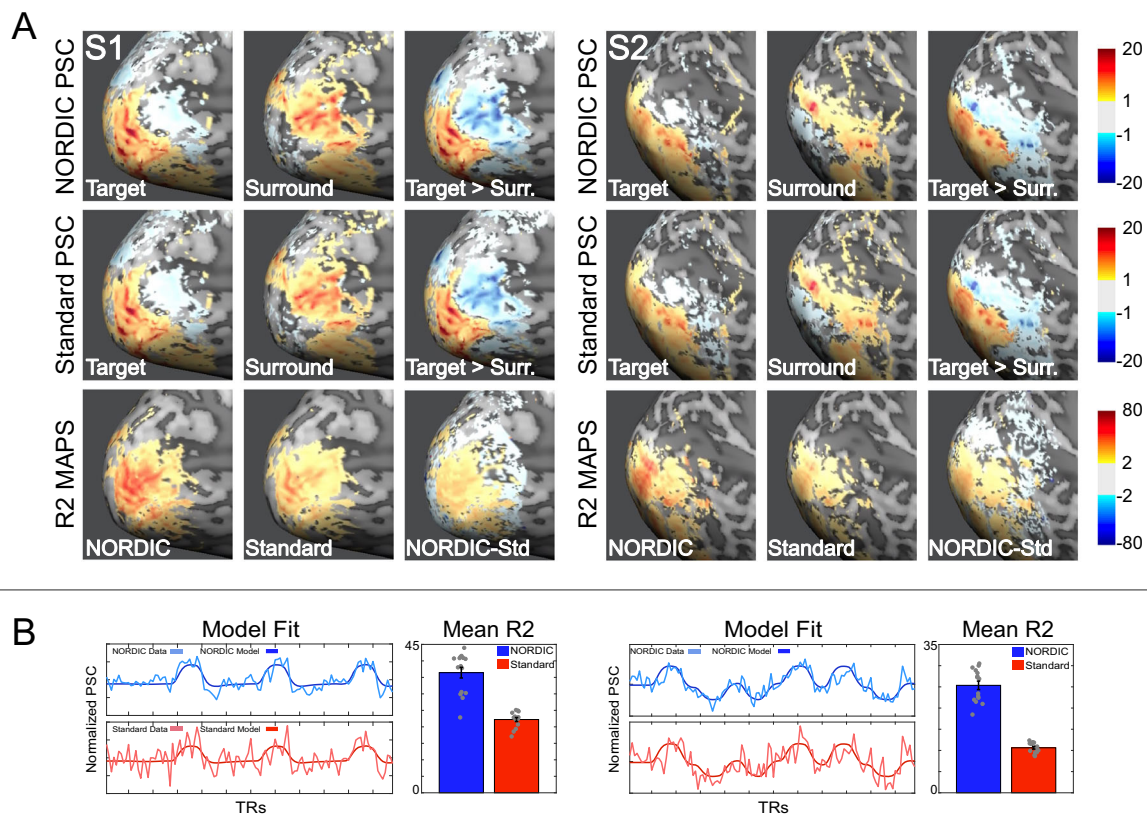
Fig. 4 PSC maps and cross-validated prediction accuracy. Panel **A**: Average percent signal change (PSC) maps and cross-validated $R2$ for both subjects S1 and S2. The top 2 rows in Panel **A** show the average (across runs) PSC maps elicited by the target (left), surround (middle) and their contrast target > surround (right), for NORDIC and Standard reconstructions, respectively. As evident by these images, PSC amplitude and the extent of stimulus-induced signal change is comparable across reconstructions (see also Supplementary Fig. 5). 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) do the $R2$ maps show higher precision of PSC estimates for NORDIC images. Panel **B**: For each subject, the left column shows 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 (2-sided paired sample *t*-test; S1: $t(14) = 12.8$, ci(11.83;16.6), Cohen's *d*: 3.31 $p < 0.01$ Bonferroni corrected; S2: $t(14) = 20.1$, ci(13.33;16.25), Cohen's *d*: 5.41 $p < .01$ Bonferroni corrected) target ROI average cross-validated $R2$ (bar plots represent the mean and error bars indicate standard errors of the mean across 15 cross-validation folds shown as gray dots). Source data are provided as source Data file.

our own implementation. We had opted to use the vendor provided reconstruction for the comparisons because we wanted to demonstrate improvements attainable with NORDIC relative to what is available for the general fMRI community, which relies on reconstruction provided by the vendor of their scanners, in this case Siemens scanners. This choice, however, raises the question of whether the gains demonstrated are not only due to NORDIC but are also partially related to differences in the reconstruction pipeline. To address this question, we have reproduced Fig. 2 for one of the four subjects (subject S2) using our offline reconstruction pipeline but *without* the NORDIC denoising step. Supplementary Fig. 15 displays the results both for "Scanner Standard", which, for ease of comparison, duplicates the functional images shown in Fig. 2, and those obtained using our offline reconstruction ("Offline Standard"). The results demonstrate that in the three cases shown (1 run, and 3 and 5 runs concatenated), the Scanner Standard and the Offline Standard produce virtually identical results, and in both cases, functional maps of 5 concatenated runs look essentially identical to a single NORDIC run.

Despite being fundamentally different from NORDIC, global SVD or PCA-based methods (e.g. refs. [25,26]) can also identify

random noise components in a time series and thus, can in principle be used to selectively suppress its contribution. Therefore, a comparison of NORDIC against such an approach would be informative. On the other hand, performing a thorough comparison under all possible conditions is beyond the scope of this paper, the primary aim of which is to introduce NORDIC and showcase its versatility (i.e. working well both in high and low SNR, cyclic and event-related paradigms, 3 and 7 T, and combinations thereof). Nevertheless, we present here the results of comparing NORDIC to two other PCA-based approaches using the 0.5 mm isotropic fMRI data where suppressing random thermal noise without incurring meaningful spatial smoothing is most challenging and at the same time, of utmost importance. Supplementary Fig. 16 illustrates these comparisons. This figure demonstrates that relative to a global PCA approach with a "white noise" criterion to identify random noise[26] (labeled PCAwn), which essentially follows an earlier SVD approach[25] with modifications, the performance of NORDIC is far superior in terms of the individual images, *t*-statistics, spatial smoothing, and the resultant functional maps. This figure also contains a comparison to a widely available implementation (named DWIdenoise) for complex and magnitude data of a relatively
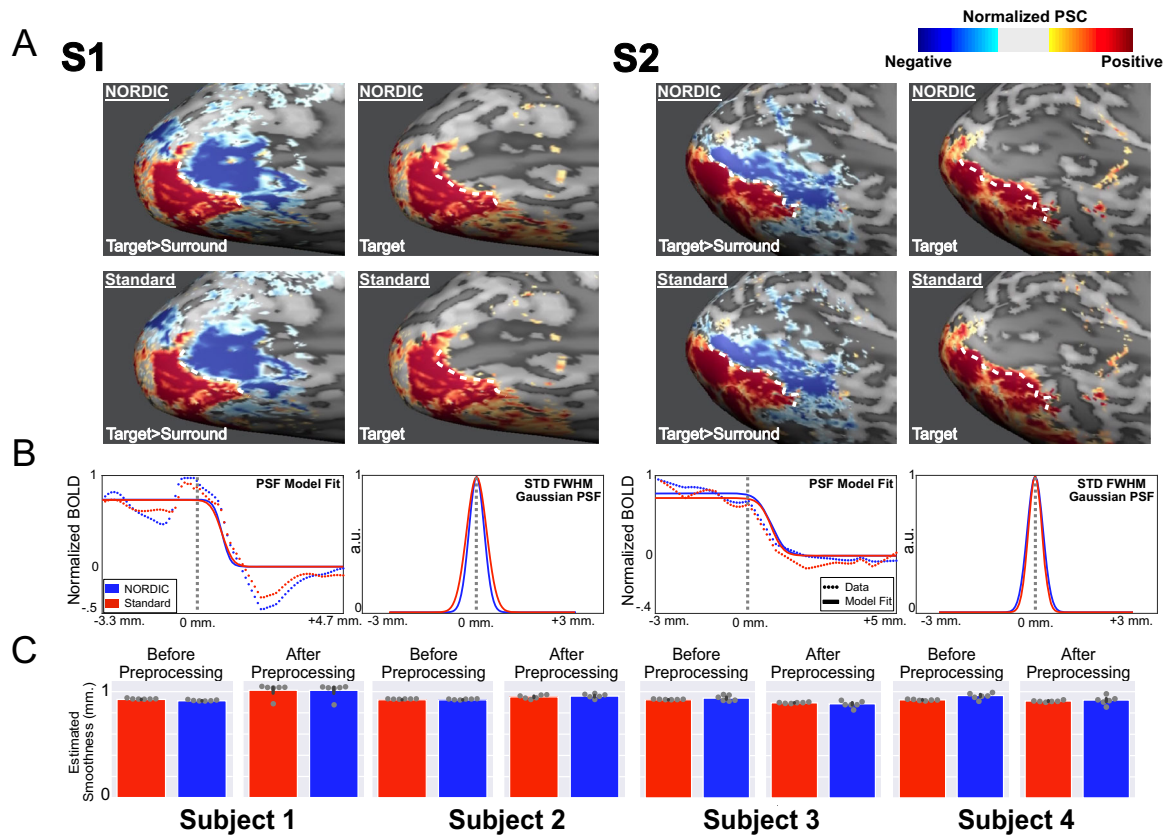
Fig. 5 Functional point spread function (PSF) and global image smoothness. Panel **A**, top row shows NORDIC normalized beta percent signal change (PSC) maps for differential mapping target (in red) > surround (in blue) (left), and the target only (right) single-condition image for subjects 1 and 2. The white dotted line is determined in the differential image as the "boundary" between the two stimulations. The same white dotted line is also superimposed on the target-only PSC map where PSC values are greater than zero but decreasing in magnitude progressively away from this "boundary" posteriorly. The functional PSF is calculated from this spread in PSC beyond the "boundary". Panel **A**, the lower row is identical to the upper row but obtained from Standard reconstruction data. Panel **B** left panel for each subject: the PSC magnitude changes (normalized to the highest value) along traces perpendicular to the "boundary" are displayed 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 gray dotted line represents the "boundary" as derived from the differential maps. Panel **B**, right panel for each subject portrays the full-width at half maximum (FWHM) standard deviation of the gaussian kernel that was convolved with a step function to model functional PSF (see the "Methods" section). Panel **C**: Mean global smoothness of images used for the fMRI time series for Standard (red) and NORDIC (blue) in four subjects, before (left panel) and after preprocessing related interpolations. Error bars represent the standard error of the mean across 6 independent runs (shown as gray dots). Source data are provided as a source Data file.

recent denoising technique called Marchenko–Pastur principle component analysis (MPPCA)[27]. Again, NORDIC outperforms this approach with respect to *t*-statistics and consequently *t*-thresholded functional maps, even though this method causes significantly larger smoothing than NORDIC.

The metrics presented in Supplementary Fig. 16 are useful in evaluating the performance of different denoising algorithms when *taken together*. However, caution should be exercised in interpreting any one metric alone. For example, the smoothness estimate for PCAwn, taken alone, suggests that this method performs relatively well. However, if we examine the EPI images, *t*-value distribution, and the related functional map (Supplementary Fig. 16), it becomes evident that this apparent preservation of spatial precision is an outcome of the failure to remove thermal noise. As explained earlier, smoothness metrics derived here utilize spatial autocorrelation, which is nonexistent for Gaussian zero-mean thermal noise. Images dominated by thermal noise would therefore show low smoothness estimates. Conversely, highly smoothed images would lead to low GLM residuals and therefore high *t*-values, albeit at the expense of degraded spatial precision.

Discussion

fMRI is inherently a low contrast-to-noise measurement where the biologically driven responses are relatively small compared to fluctuations (i.e. "noise") in the amplitude of the signal in the fMRI time series. Certainly, the thermal noise of the MR detection[10,11] contributes to this "noise". Physiological processes of respiration and cardiac pulsation[15,28–30], and, for taks and stimulus fMRI, the spatially correlated spontaneous fluctuations ascribed to functional networks in rsfMRI[31] represent other sources of tSNR degradation, which, unlike thermal noise, are non-white in nature[16–19]. These non-Gaussian sources of signal fluctuations are proportional to signal magnitude[14,32–35]; as such, they become dominant only when a voxel's signal (which is proportional to voxel volume) is large compared to instrumental thermal noise, as encountered, for example, with low spatial resolutions, high flip angles used in conjunction with long TRs, and at high magnetic fields[13,14,36,37].

Reliably detecting the relatively weak biologically driven responses in the presence of the afore-mentioned noise contributions requires significant efforts to clean up the fMRI time series. This problem was addressed as early as approximately two decades ago using component analysis based on SVD[25], and
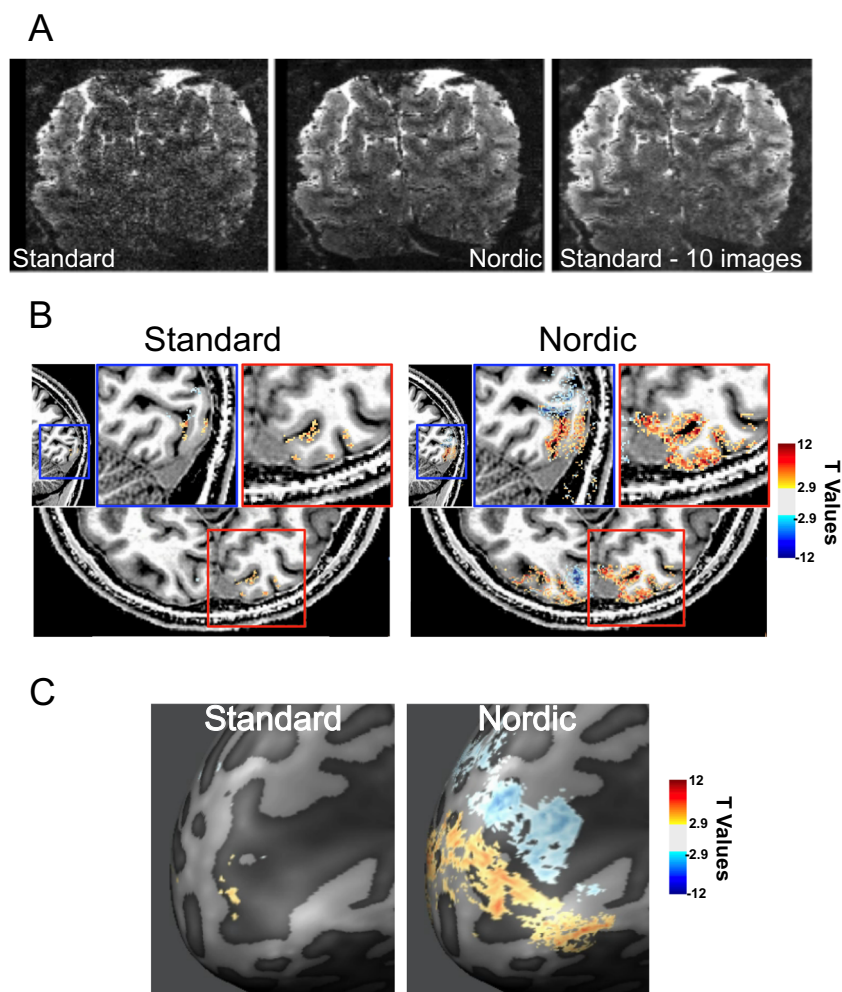
Fig. 6 3D GE EPI images and fMRI data obtained with 0.5 mm isotropic voxels. Panel **A** shows a single slice from a single time point in the consecutively acquired volumes forming the fMRI time series for Standard (left) and NORDIC (middle) images. The right panel shows the average of 10 images of the same slice for the Standard reconstruction. Panel **B** shows *t*-thresholded (|*t*| ≥ 2.9) functional maps (for the contrast target > surround on a T_1 weighted anatomical image for standard (left) and NORDIC (right) reconstructions for a saggital and axial slice (with related zoom-ins on the sagittal (blue) and axial (red) planes). Panel **C** shows the same *t*-maps as in panel **C** on the inflated cortical space and at different *t*-thresholds. No spatial smoothing or masking was applied to the data.

subsequently, PCA and ICA[26] to decompose the fMRI time series into components containing task/stimulus-response, structured noise, and thermal (random) noise. Although these early holistic approaches have not been widely adopted, numerous methods using PCA and ICA components analysis in various ways have subsequently been introduced and employed almost exclusively on the suppression of the non-white confounds (e.g. refs. [16–19,38,39] and references therein). In this paper, we introduce a method named NORDIC aimed at improving the detectability of the inherently small fMRI signals by selectively targeting the suppression of thermal noise. As such, the approach represents a change in direction that shifts the focus from structured, non-white noise to thermal noise, leaving the suppression of the structured noise, if desired, to a subsequent complementary step.

NORDIC is fundamentally different in its approach to the above-referenced PCA and ICA methods. Although these previous methods for the most part have concentrated on identifying structured noise[17], some of them also provide a strategy to selectively suppress thermal noise; they do so using a global PCA analysis and an empirical threshold for the differentiation of noise and signal components, in some cases working best in the

presence of a clear periodic temporal signature in the signal[26], which naturally limits their general utility. In contrast, NORDIC uses a local (patch) approach, experimental characterization of thermal noise independent of the functional imaging data, and a well-defined objective principle to identify the threshold for its suppression. Especially for low SNR, high-resolution fMRI data, a global component analysis may be suboptimal (see the comparison in Supplementary Fig. 16); as such, in such data where the need for thermal noise suppression is immense, spatial smoothing has been the method of choice to improve SNR and fCNR even at the risk of degrading spatial specificity. In contrast, our results demonstrate that NORDIC is particularly (but not exclusively) useful for such low SNR high-resolution fMRI data (Fig. 6, Supplementary Figs. 14, 16).

NORDIC and its application in diffusion-weighted imaging (dMRI) was previously described[40] and was shown to yield superior results to the recently introduced MPPCA[27] method, which also selectively targets thermal noise removal. It is difficult to precisely identify the components that are removed in MPPCA, although its application leads to better results in dMRI[27,40] and increased reproducibility in rsfMRI[41,42]. In contrast, NORDIC yields a parameter-free threshold, correlated with

the global thermal noise level, to remove signal components that cannot be distinguished from i.i.d, zero-mean Gaussian data, which is attributable to thermal noise. Even though the remaining signal components also contain some residual thermal noise (see discussion in the "Methods" section), the overall impact is a significant improvement in tSNR for NORDIC compared to Standard data (Fig. 1C and Supplementary Figs. 2 and 3) as well as to MPPCA (Supplementary Fig. 16).

Difference of NORDIC vs. Standard images show only noise, which, when averaged over all the images in the fMRI times series demonstrates equivalence to the g-factor maps (Supplementary Fig. 1), without evidence of edge effects or features of the imaged object; additionally, the FFT power spectra (Supplementary Fig. 7) display only a broadband decrease in the magnitude of the spectrum without impacting the various peaks detected at specific frequencies associated with the stimulus presentation or physiologic fluctuations. These observations are consistent with the expectation that NORDIC suppresses random noise associated with the thermal noise of the MR measurement without perturbing the image.

t-Values are a useful metric in evaluating functional mapping studies. Denoising algorithms inherently alter the dimensionality of the data and, consequently, the 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[43]. Here we do not attempt to address this issue, which is beyond the scope of this work as it relates not only to denoised time-series but is intrinsic to fMRI in general. We chose instead to compute our t-values using Eq. (2) (see the "Methods" section) to provide a measure of activation relative to GLM residual noise. Thus, our activation maps are based on t-value rather than p-value thresholds, although we give the equivalent p value as a reference for the Standard reconstruction.

At the same t-threshold, the extent of voxels showing stimulus-invoked signal changes that pass the t-threshold is considerably larger for the NORDIC processed single run (Figs. 2 and 3; Supplementary Figs. 4, 8–12) and equivalent to activation maps produced by concatenating 3–5 runs of the Standard data. This was also consistently observed for 3 and 7 T data obtained with different resolutions, paradigms, and cortical regions (Supplementary Figs. 8–19, 12). These observations are expected given the fact that NORDIC improves more than 2-fold the trial-to-trial precision of single-voxel PSC estimates while not impacting the magnitude of the PSC (Figs. 3 and 4). Thus, NORDIC better estimates the stimulus-evoked responses and does so 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 NORDIC is its ability to preserve spatial precision of the individual images of the fMRI time series, as well as the precision of the functional response. Thermal noise associated with the MR process can and often is suppressed with spatial filtering, which smooths (i.e. blurs) the images, increasing the SNR and consequently the tSNR[44]; this improves the t-values (Supplementary Figs. 8 and 11) and also, when applied with a Gaussian kernel, serves the purpose of making more valid the assumption of smoothness for FWER control based on random field theory (RFT) approaches widely used in the fMRI community. However, the resultant spatial blurring leads to an undesirable loss of spatial precision. NORDIC, on the other hand, suppresses thermal noise and has the same impact on t-values as spatial-smoothing (Supplementary Figs. 8 and 11) but *without* spatial blurring of either the individual images themselves (Figs. 5C, 6, and Supplementary Fig. 14 and also see discussion in the "Methods" section) or the

functional PSF estimates in the visual cortex (Fig. 5A, B), yielding PSF values consistent with previous reports[22,45].

NORDIC can be said to improve the spatial specificity to neuronal activity changes by reducing false positives, and negatives. However, there could be additional benefits in specificity due to the ultrahigh resolutions enabled by NORDIC. At sufficiently high enough resolutions, the draining vein confound (e.g. see ref. [46]) of GE BOLD fMRI is less of a problem because partial voluming and spatial averaging will be less and there would exist many voxels unaffected by this confound providing access to tissue responses, just like in optical imaging with intrinsic signals where blood vessels are visible but the high resolution permits visualization of the tissue responses in between blood vessels. There is, however, an additional advantage that can arise from the small voxel sizes achievable with NORDIC. GE BOLD fMRI is based on the voxel-wise measurement of signal *amplitude* after it is allowed to decay for an echo time TE with the rate constant $1/T_2^*$. T_2^* strongly depends on intravoxel B_0 inhomogeneities, hence on neuronal activity because of the extravascular B_0 gradients generated by deoxyhemoglobin-containing blood vessels. However, in the limit voxel dimensions become small compared to the spatial scale of these extravascular B_0 gradients, intravoxel inhomogeneities, hence their contribution to $1/T_2^*$, also become small, reducing the detectability of stimulus/task-induced alterations as an *amplitude* change in GE BOLD fMRI. For a given blood vessel, this limit is determined approximately by $(\delta d/r_b)$ where δd is the voxel dimension and r_b is the blood vessel radius, and the distance from the blood vessel since the gradient becomes rapidly shallower with increasing distance from the blood vessel. Thus, high resolutions enabled by NORDIC and other advances will lead to an intrinsic shrinkage of the spatial extent of the draining vein confound in GE BOLD fMRI and ultimately its suppression. Extravascular B_0 gradients still exist, of course. However, in this limit, they will show up as a phase difference among the different voxels. Such phase effects mixed with amplitude changes were already reported and used to account for large draining confound in GE BOLD fMRI[47]. As the resolutions increase, however, the amplitude effects will become smaller, leaving behind ultimately only the phase perturbation. The intravascular BOLD effect will still persist and will be a source of unwanted BOLD signals at lower magnetic fields like 3 T[48] but not at ≳7 T where the very short T_2 of blood assures its elimination[49].

In the mammalian cortex there exist elementary cortical units of operation, consisting of several hundreds or thousands of neurons, that are spatially clustered and repeated numerous times in each cortical area. These mesoscopic scale ensembles are the focus of extensive research carried out in animal models by invasive techniques, such as optical imaging or electrophysiology. However, these techniques cannot be used in human studies because of their invasive nature. Therefore, the ability to generate functional maps at the level of these elementary units by MR methods is critically important and has been shown to be feasible[7–9]. However, current achievable resolutions (~0.8 mm isotropic or non-isotropic voxels of equal volume) and the responses detected at such high resolutions are at best marginal. Overcoming this barrier with reasonable acquisition times has not been possible. For example, it has been possible to detect axis of motion features in the human MT[50] but not the *direction of motion* subclusters that distinguish the motion in the two different directions along a given axis. It has been possible to demonstrate layer specific activations aimed at studying laminar organization but not with sufficient resolution to even distinguish three layers across the cortex without partial voluming and with Nyquist sampling; at least three (ideally more) distinct layers are required in order to clearly differentiate feedforward inputs arriving primarily into layer 4, local computations and cortico-

cortical inputs shaping responses in layers 2/3, and outputs to
other brain areas from layers 2/3, and 5/6.

The afore-described barrier and its limitations on neu-
roscientific research were recognized in the first report of the
BRAIN Initiative Working group[6,51], which challenged the MR
community to overcome it and achieve whole brain imaging
studies with at least 0.1 μL voxel volumes (e.g. 0.46 or ~0.5 mm
isotropic resolution). We demonstrate here that this goal is
achieved with NORDIC (Fig. 6) at 7 T and likely will soon be
surpassed when multiplicative gains will be attained combining
NORDIC with additional independent gains from acquisition
methods, higher magnetic fields[52], high channel count RF coils
employed synergistically with very high magnetic fields[53,54] and
image reconstructions methods (e.g. refs. [55,56]).

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 7 T
mapping at high spatial resolution, as well as for 3 and 7 T 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 struc-
tured, non-white noise removal. The cumulative gains are
expected to bring in transformative improvements in fMRI,
permitting higher resolutions at 3, 7 T and higher magnetic fields,
more precise quantification of functional responses, faster
acquisitions rates, significantly shorter scan times, and the ability
to reach finer scale mesoscopic organizations that have been
unreachable to date.

Methods
Image reconstruction
2D slice selective accelerated acquisitions. For 2D acquisition with phase-encoding
undersampling and/or SMS/MB acquisition, the GRAPPA and slice-GRAPPA
reconstructions were used as outlined in ref. [57]. A single kernel G_j^{ch} is constructed
for SMS/MB with/without phase-encoding undersampling such that for each slice,
j, and channel, ch,

$$G_j^{ch}(S_{MB}) = SB_j^{ch} \forall j, ch \tag{1}$$

where S_{MB} denotes the acquired SMS/MB k-space, and SB_j^{ch} denotes the recon-
structed k-space for the slice j and channel ch. The kernels G_j^{ch} are calculated
similarly as in unbiased slice-GRAPPA from the measured individual slices SB_i
with $S_{MB} = \sum_{i=1}^{MB} SB_j$.

3D accelerated acquisitions. For 3D acquisitions with phase-encoding under-
sampling only, a gradient recalled echo (GRE) based Nyquist-sampled auto-cali-
bration signal (ACS) reference acquired without slice-phase-encoding (a single
slice-phase-encoding plane) was used. A Fourier transform was first applied along
the slice-phase-encoding, and then k-space interpolation along the phase-encoding
direction was performed with GRAPPA-weight calculated from the ACS reference.

g-factor noise for image-reconstruction. g-factors were calculated building on the
approach outlined in ref. [58] for g-factor quantification in GRAPPA reconstructions
and detailed in ref. [57]. The same ESPIRIT sensitivity profiles used for image
reconstructions were also used for the determination of the quantitative g-factor.

NORDIC PCA. Let $m(\mathbf{r}, t) \in \mathbb{C}^{I_1 \times I_2 \times I_3 \times Q}$ denote a complex-valued volumetric
fMRI image series following an accelerated parallel imaging acquisition, where Q is
the number of temporal samples and I_1, I_2, I_3 the matrix size of the volume. The
flow chart in Fig. 7, adapted from ref. [40], illustrate the principles of NORDIC
denoising of this dataset $m(\mathbf{r}, t)$ and the details of the noise model, locally low rank
model, threshold selection, and patch averaging.

Noise model. Images in MRI are inherently complex-valued but constructed as real-
valued by using the magnitude of the images. This transformation changes the
thermal Gaussian i.i.d. noise in the original measurement to be Rician for mag-
nitude of coil-combined images or non-central Chi[2] distributed when combining

multiple magnitude images from different coils. Furthermore with parallel imaging
reconstruction, the noise undergoes a spatially varying amplification, which is
characterized by the geometry-factor, $g(\mathbf{r})$. In NORDIC, a signal and noise scaling
is performed on the complex valued data as $m(\mathbf{r}, t)/g(\mathbf{r})$ to ensure zero-mean and
spatially identical noise in a given patch (left-most column, Fig. 7). For NORDIC
processing, a sensitivity weighted channel combination[59] is applied to the accel-
erated dataset[57] to maintain complex-valued Gaussian noise[60] of the combined
image, and the images are transformed to magnitude images only after denoising.

Locally low rank model. For locally low rank (LLR) processing, a fixed $k_1 \times k_2 \times k_3$
patch is extracted from each volume in the series, and the voxels in each patch from
each volume is vectorized as $\mathbf{y}_t$, to construct a Casorati matrix $\mathbf{Y} =$
$[\mathbf{y}_1, \cdots, \mathbf{y}_t, \cdots, \mathbf{y}_Q] \in \mathbb{C}^{M \times Q}$ with $M = k_1 \times k_2 \times k_3$, and Q representing the
number of volumes (time points) in the fMRI time series. The concept of NORDIC
is to estimate the underlying matrix $\mathbf{X}$ in the model where $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, and
$\mathbf{N} \in \mathbb{C}^{M \times Q}$, where $\mathbf{N}$ is additive Gaussian noise.

LLR modeling assumes that the underlying data matrix $\mathbf{X}$ has a low-rank
representation. For NORDIC, $k_1 \times k_2 \times k_3$ is selected to be a sufficiently small patch
size so that no two voxels within the patch are unaliased from the same acquired
data for the given acceleration rate[40], ensuring that the noise in the pixels of the
patch are all independent. LLR methods typically implement the low-rank
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, $\lambda(j), j \in \{1, \ldots, N\}$. Then the singular
values below a threshold $\lambda(j) < \lambda_{thr}$ are changed 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 $\mathbf{Y}_L = \mathbf{U} \cdot \mathbf{S}_{\lambda_{thr}} \cdot \mathbf{V}^H$.

Hyperparameter selection. While the threshold in NORDIC is chosen automatically
without any empirical tuning, the method itself has hyperparameters related to the
patch size that determine the size of Casorati matrices. In NORDIC, $k_1 \times k_2 \times k_3$ is
selected with $M \approx 11 \cdot Q$, and $k_1 = k_2 = k_3$, as determined heuristically in Moeller
et al.[40]. We note that the choice of patch size with a $M : Q$ ratio of 11:1, can be
more challenging to accommodate for long fMRI runs since Q, is the number of
samples in the time series, especially in light of the requirement that no two voxels
within a patch are unaliased from the same acquired data. For whole brain rsfMRI,
as in the HCP, for example, $M \approx 11 \cdot Q$ can be maintained. If there is an issue
fulfilling this requirement, the geometry of the patch may be adjusted to something
different than $k_1 = k_2 = k_3$.

The patches can be either 2D or 3D, and while 2D patches may better fit with
the temporal dynamics of the acquisition, the data independence constraint of no
two voxels within the patch being unaliased from the same acquired data can be
challenging. For longer series, the constraint of $M \approx 11 \cdot Q$ may either in itself not
be satisfied simultaneously with the data independence, or it may be further
difficult in the presence of phase-encoding ghosting e.g. from fat or eddy currents.
3D patches are less limited in this regard and also better capture spatially similar
signals.

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. The
threshold that ensures the removal of components that are indistinguishable from
Gaussian noise is the largest singular value of the noise matrix $\mathbf{N}$. While this
threshold is asymptotically specified through the Marchenko–Pastur distribution,
for practical finite matrix sizes, we numerically estimate this value via a Monte-
Carlo simulation[40]. To this end, random matrices of size $M \times Q$ are generated with
i.i.d. zero-mean entries, whose variance match the experimentally measured ther-
mal noise, σ^2, in $\mathbf{Y}$. The thermal noise level can be determined after g-factor
normalization from an appended acquisition without RF excitation (a noise
acquisition) to the series, or from a region of interest outside the brain devoid of
signal contributions or it can be determined from a receiver noise pre-whitening
acquisition. In this paper the first method of utilizing an additional noise-
acquisition has been employed. Then the empirical mean value of the largest
singular value is used as the numerical threshold.

The degree of noise removal. Though NORDIC removes zero-mean, i.i.d. Gaussian
noise, it does not remove all of it. This can be explained more formally by con-
sidering one of the $M \times Q$ Casorati matrices we are trying to denoise based on the
model $\mathbf{Y} = \mathbf{X} + \mathbf{N}$. According to our model, $\mathbf{Y}$ is the observed noisy data, $\mathbf{N}$ is a
matrix whose entries are zero-mean, i.i.d. Gaussian, and $\mathbf{X}$ is the low-rank data
matrix. More concretely, the low-rank condition states $rank(\mathbf{X}) = r \ll \min\{M,
Q\} = Q$ (latter equality due to our choice of M). For ease of explanation, also
assume that all non-zero singular values of $\mathbf{X}$ are sufficiently above the noise level.
Thus, when the singular value decomposition of $\mathbf{Y}$ is performed, it will have r
singular values that contain a combination of signal component from $\mathbf{X}$ and noise
component from $\mathbf{N}$, while the remaining $(Q-r)$ singular values will only have
contributions from noise $\mathbf{N}$. Since the thresholding is performed at the level of the
largest singular value of the noise matrix, NORDIC will remove the noise from all
these $(Q-r)$ components, as they cannot be distinguished from zero-mean
i.i.d. Gaussian noise (i.e. random noise). On the other hand, the r singular values
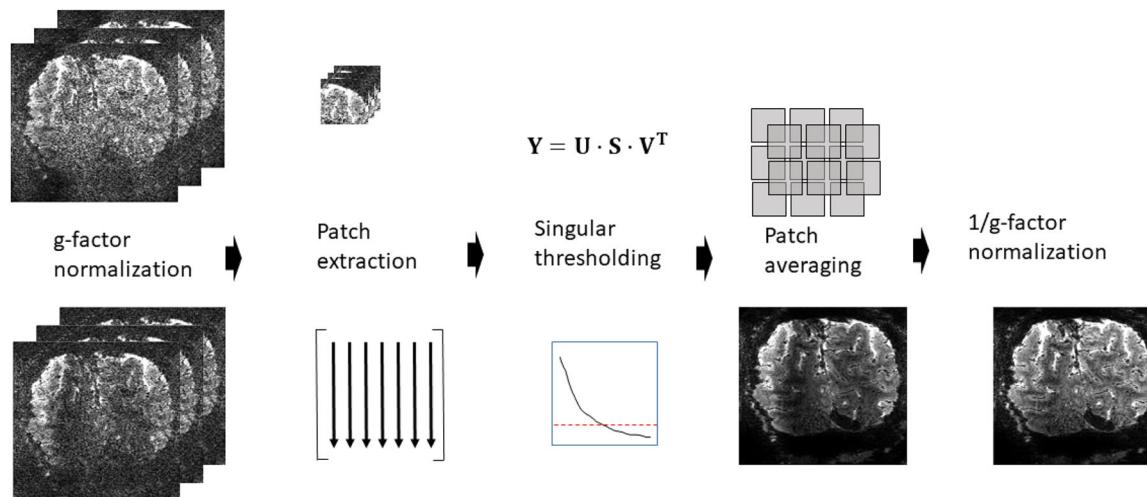
Fig. 7 Flowchart of the NORDIC algorithm for a series $m(\mathbf{r}, \tau)$. First, to ensure i.i.d. noise the series is normalized with the calculated g-factor kernels as $m(\mathbf{r}, \tau)/g(\mathbf{r})$. From the normalized series, the Casorati matrix $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_j, \cdots, \mathbf{y_N}]$ is formed, where $\mathbf{y}_j$ is a column vector that contains the voxel values in each patch. The low-rank estimate of $\mathbf{Y}$ is calculated as $\mathbf{Y}_L = \mathbf{U} \cdot \mathbf{S}_{\lambda_{\text{thr}}} \cdot \mathbf{V}^T$, where the singular values in $\mathbf{S}$, $\lambda(i)$ are set to 0 if $\lambda(i) < \lambda_{\text{thr}}$. After re-forming the series $m^{LLR}(\mathbf{r}, \tau)$ with patch averaging, the normalization with the calculated g-factor is reversed as $m^{NORDIC}(\mathbf{r}, \tau) = m^{LLR}(\mathbf{r}, \tau) \cdot g(\mathbf{r})$.

that are above the threshold will be unaffected by NORDIC processing. However, these r singular values have contributions from both noise and signal components, though these components will be dominated by the signal. Thus, the final denoised estimate generated from these singular values and their corresponding singular vectors will have residual Gaussian noise in them. Since $r \ll Q$ due to low-rank assumption, majority of the thermal noise is removed by virtue of thresholding $(Q–r)$ singular values, but a small amount of thermal noise that are on the remaining r singular components will remain in the final estimate. As a side note, this remaining thermal noise will be further reduced due to patch averaging in processing, but this effect is difficult to quantify.

Patch averaging. The patches arising from these thresholded Casorati matrices are combined by averaging[61] overlapping patches to generate the denoised image series $m^{LLR}(\mathbf{r}, \tau)$. The averaging of patches can be performed with patches having different geometries, i.e. k_1, k_2, k_3, and the averaging can be identically weighted or weighted by the number of non-zero λ's. In NORDIC for fMRI, direct averaging with identical weights is used, similar to the previous use of NORDIC in dMRI, where it was shown that there was no difference from using weighted averaging[40]. The patch-averaging is itself a denoising step[62] which reduces the residual contributions of noise. In NORDIC for fMRI, with typically $Q>100$, and $M>1000$, we used patch averaging with 25–50% overlap, and the difference between this and using all combinations of patches was minimal, but led to substantial savings in computational time.

Finally, to obtain $m^{NORDIC}(\mathbf{r}, t)$ the denoised volumes $m^{LLR}(\mathbf{r}, t)$ are multiplied back with the g-factor map $g(\mathbf{r})$ to correct the signal intensities.

Spatial blurring. It may seem counter-intuitive that noise can be removed without introducing spatial blurring. The main idea behind the locally low-rank decomposition is to separate out the noisy Casorati matrix $\mathbf{Y}$ into two components as $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, where $\mathbf{X}$ is assumed to be low-rank, and $\mathbf{N}$ is Gaussian noise. Then the algorithm thresholds to remove all principal components of $\mathbf{Y}$, whose singular values are below the threshold that is automatically determined in NORDIC by the noise level. This will remove both contributions from $\mathbf{N}$ and from $\mathbf{X}$. This is analogous to the concept of image compression, where part of the data is removed (e.g. some of the DCT coefficients in JPEG compression), but the end result is visually indistinguishable from the uncompressed image, as long as the compression level is not too high. In this analogy, the compression is done via removing some of the components of the low-rank $\mathbf{X}$, but due to its low-rank property, this does not fundamentally alter its visualization. Additionally, the compression level in conventional image compression is analogous to the SNR/threshold level in our method. A numerical simulation of the threshold and patch size relative to zero-mean Gaussian noise was performed in Moeller et al.[40].

Participants. To test the impact of NORDIC on fMRI, we acquired 10 data sets on four (2 females) healthy right-handed subjects (age range: 27–33), with different stimulation paradigms, acquisition parameters and field strengths (see the sections "Stimuli and procedure" and "MRI imaging acquisition and processing" sections). All subjects had normal, or corrected vision and provided written informed consent. The study complied with all relevant ethical regulations for work with human

participants. The local IRB at the University of Minnesota approved the experiments.

Stimuli and procedure. We tested the impact of NORDIC on fMRI across four experimental paradigms:

1. Block design visual stimulation
2. Fast event related visual stimulation design
3. Fast event related auditory stimulation design
4. Resting state.

Block design visual stimulation. We implemented standard block design visual stimulation paradigms (see Fig. 1A) for four acquisition types. These included the two 3 T fMRI studies, the 0.8 mm isotropic resolution 7 T fMRI and the 7 T 0.5 mm isotropic resolution fMRI datasets (see the section "MRI imaging acquisition and processing" section). The experimental procedure consisted of a standard 12 s on, 12 s off for the 7 T 0.8 mm isotropic voxel acquisitions, and for the 3 T datasets, and a 24 s on, 24 s off for the 7 T 0.5 mm isotropic voxel acquisitions (see Fig. 1A). The difference in block length between the two resolutions was implemented to account the difference in volume acquisition time between the 0.8 mm iso (i.e. volume acquisition time = 1350 ms) and the 0.5 mm iso acquisitions (i.e. volume acquisition time = 3652 ms). The stimuli consisted of a center (i.e. target) and a surround square checkerboard counterphase flickering (at 6 Hz) gratings (Fig. 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 7 T 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 through a mirror placed in the head coil.

Each run lasted just over two and a half minutes for the 0.8 mm 7 T and the 3 T acquisitions (i.e. 118 volumes at 1350 ms TR) and just over 5 min for the 0.5 mm 7 T acquisitions (85 volumes at 3654 ms volume acquisition time), beginning and ending with a 12 or 24 s red fixation dot centered on a gray background. Within each run, each visual condition, target and surround, was presented three times. For the 0.5 mm iso data sets, we collected eight experimental runs; for the 0.8 mm iso 7 T and the two 3 T data sets, participants underwent eight runs, two 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. For the 0.8 mm 7 T acquisition on S3, run 8 had to be discarded due to excessive movement.

Fast event-related visual design. The visual fast event related design consisted of six runs of a face detection task, with a 2 s on, 2 s off acquisition. Each run lasted approximately 3 min and 22 s and began and ended with a 12 s fixation period. Importantly, we introduced 10% blank trials (i.e. 4 s of fixation period) randomly interspersed amongst the images, effectively jittering the ISI. Stimulus presentation was pseudorandomized across runs, with the only constraint being the non-occurrence of two consecutive presentations of the same phase coherence level. Behavioral metrics, including reaction time and responses to face stimuli

indicating participants' perceptual judgments (i.e. face or no face) were also recorded.

We used grayscale images of faces (20 male and 20 female). We manipulated the phase coherence of each face, from 0% to 40% in steps of 10%, resulting in 200 images (5 visual conditions × 20 identities × 2 genders). We equated the amplitude spectrum across all images. Stimuli approximately subtended 9 degrees of visual angle. Faces were cropped to remove external features by centering an elliptical window with uniform gray background to the original images. The *y* diameter of the ellipse spanned the full vertical extent of the face stimuli and the *x* diameter spanned 80% of the horizontal extent. Before applying the elliptical window to all face images, we smoothed the edge of the ellipse by convolving with an average filter (constructed using the "fspecial" function with "average" option in MATLAB. This procedure was implemented to prevent participants from performing edge detection, rather than the face detection task, by reacting to the easily identifiable presence of hard edges in the face images.

Fast event-related auditory design. Stimuli consisted of sequences consisting of four tones. For each sequence, tones were presented for 100 ms with a 400 ms gap in between them (sequence duration 1.6 s). The sequences were presented concomitantly with the scanner noise (i.e. no silent gap for sound presentation was used) and 36 tone sequences were presented in each run, a session consisted of 10 runs of about 6 min each. Tone sequences were presented following a slow-event related design with an average interval of 6 TR's (ranging between 5 and 7 TR's, TR = 1.6 s)."

Resting state. The resting state acquisition consisted of four 10 min runs. Data were obtained at 3 T with 3 T HCP acquisition parameters (see section below). No stimulus presentation occurred and participants were instructed to stay still, minimize movements and fixate on a visible crosshair.

MR imaging acquisition and processing

7 T Acquisition parameters. All 7 T functional MRI data were collected with a 7 T Siemens Magnetom System with a single transmit and 32-channel receive NOVA head coil.

We collected four variants of T2*-weighted images with different acquisition parameters, tailored to the different experimental needs. Specifically, for *block design* visual stimulus paradigm at 7 T we collected 0.5 mm iso voxel (T2*-weighted 3D GE EPI, single slab, 40 slices, TR 83 ms, Volume Acquisition Time 3654 ms, 3-fold in-plane undersampling along the phase encode direction, 6/8ths in plane Partial Fourier, 0.5 mm isotropic nominal resolution, TE 32.4 ms, Flip Angle 13°, Bandwidth 820 Hz). The 0.8 mm iso voxel acquisition used T2*-weighted 2D GE SMS/MB EPI, 40 slices, TR 1350 ms, MB factor 2, 3-fold in-plane undersampling along the phase encode direction, 6/8ths Partial Fourier, 0.8 mm isotropic nominal resolution, TE 26.4 ms, flip Angle 58°, Bandwidth 1190 Hz. For the *auditory event related design*, we used a comparable submillimeter acquisition protocol (2D GE SMS/MB EPI 42 slices, TR 1600 ms, MB factor 2, 3-fold in-plane undersampling along the phase encode direction, 6/8ths Partial Fourier, 0.8 mm isotropic nominal resolution, TE 26.4 ms, Flip Angle 61°, Bandwidth 1190 Hz)

For the *visual fast event-related design*, we used the 7 T HCP acquisition protocol (2D GE SMS/MB EPI, 85 slices TR 1 s, MB factor 5, 2-fold in-plane undersampling along the phase encode direction, 7/8ths Partial Fourier, 1.6 mm isotropic nominal resolution, TE 22.2 ms, Flip Angle 51°, Bandwidth 1923 Hz).

3 T Acquisition parameters

We recorded data employed the *block design visual stimulus* paradigm using two sequences varying in resolution: Acquisition sequence 1 used the 3 T HCP protocol parameters (72 slices, TR = 0.8 s, MB = 8, no in-plane undersampling 2 mm isotropic, TE = 37 ms, flip angle = 52°, bandwidth = 2290 Hz/pixel). Acquisition sequence 2 parameters were 100 slices, TR = 2.1 s, MB = 4, in-plane undersampling factor = 2, 7/8 partial Fourier, 1.2 mm isotropic, TE = 32.6 ms, flip angle = 78°, bandwidth = 1595 Hz/pixel.

For the resting state data we used the acquisition sequence 1 detailed above (i.e. the 3 T HCP protocol).

For all acquisitions, flip angles were optimized to maximize the signal across the brain for the given TR. For each participant, shimming to improve B_0 homogeneity over occipital regions was conducted manually.

T1-weighted anatomical images were obtained on a 3 T Siemens Magnetom Prisma$^{\text{fit}}$ system using an MPRAGE sequence (192 slices; TR, 1900 ms; FOV, 256 × 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 gray 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

All 7 T functional data preprocessing was performed in BrainVoyager. Preprocessing was kept at a minimum and constant across reconstructions. Specifically, we performed slice scan timing corrections for the 2D data (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 second-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.

3 T dicom files were converted using dcm2niix[63]. All subsequent 3 T functional data preprocessing was performed in AFNI version 19.2.10[64]. Conventional processing steps were used, including despiking, slice timing correction, motion correction, and alignment to each participant's anatomical image.

EPI data were aligned to T_1-weighted images. For all MB data sets (i.e. all acquisitions other than the 3D 0.5 mm iso images), anatomical alignment was performed on the Single Band Reference (SBRef) image which was acquired to calibrate coil sensitively profiles prior to the MB acquisition and has no slice acceleration or T_1-saturation, yielding higher contrast[65].

GLMs and tSNR. Stimulus-evoked functional maps were computed in Brain-Voyager for all 7 T datasets and in AFNI for the 3 T datasets. ROI definition and contrast maps were also computed using these software. Subsequent analyses (i.e. ROI based and functional point spread function measurements) were performed in MatLab using a set of tools developed inhouse.

Temporal 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 the section "NORDIC vs. Standard statistical analyses" section below) of the responses evoked by the target condition for each voxel within the left retinotopic representation of the target in V1 (see below).

ROI 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. While this approach may overinflate statistical power and misrepresent the size of the ROI, it also ensures identical ROIs across reconstructions, which was the main goal in this case. GLM *t*-values can be thought of as beta estimates divided by GLM standard error according to this equation:

$$t = c\prime b / \sqrt{\text{Var}(e)c\prime(X\prime X)^{-1}c} \qquad (2)$$

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 ($p < 0.05$ Bonferroni corrected) to define the left hemisphere retinotopic representation of the target stimulus within the gray 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 PSC beta weights and the related *t*-values for 2–6 runs. For each *n*-run GLM, we computed independent GLMs for all possible run combinations (see the section "Comparing extent of activation" section 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 described earlier. 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[23]) 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 statistically significant differences in stimulus-evoked BOLD amplitudes and noise levels across reconstruction algorithms, we compared the ROI voxel mean PSC 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:

$$Data \sim Cond + (1|runs) + (1|trials) \tag{3}$$

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:

$$R2 = 1\text{-}SS_{error}/SS_{total} \tag{4}$$

and, in our specific case

$$SS_{error} = \sum (f(x) - x)^2 \tag{5}$$

$$SS_{total} = \sum S(x_{(i)} - \mu(x))^2 \tag{6}$$

In Eqs. (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 $R2$s 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 PSC 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 10,000 times, leading to 10,000 means. We sorted these 10,000 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 $p < 0.05$ 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 1–6 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[23]. This procedure not only ensured sampling from all runs, but it also decreased the impact of extreme values[23].

Quantifying BOLD images smoothness. Global smoothness estimates from each reconstruction prior to preprocessing ('pre') and following all data preprocessing, just prior to the GLM ('post'). This was performed using 3dFWHMx from AFNI[64] 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 auto-correlation 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 Shmuel et al.[22]. We estimated the BOLD functional PSF on all individual runs independently for the Standard and NORDIC reconstructions. In brief the analysis was implemented as follows: We first identified the anterior most retinotopic representation of the target's edge in V1 separately in Standard and NORDIC reconstructed data. This was achieved by computing the contrast target > surround on all runs concatenated within each group (Standard vs. NORDIC) and identifying those voxels showing differential BOLD closest of 0 (Fig. 5). 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. Independently per run, we then drew 10 traces orthogonal to the retinotopically anterior most edge of the target. We estimated the BOLD functional PSF on all individual runs independently for the Standard and NORDIC reconstructions. In brief the analysis was implemented as follows: We first identified the anterior most retinotopic representation of the target's edge in V1 separately in Standard and NORDIC reconstructed data. This was achieved by computing the contrast target > surround on all runs concatenated within each group (Standard vs. NORDIC) and identifying those voxels showing differential BOLD closest of 0 (Fig. 5). 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. Independently per run, we drew 10 traces orthogonal to the retinotopically anterior most edge of the target. We then 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 consisting of a step-function (representing infinitely precise PSF) convolved with a gaussian[22] with three free parameters. The three parameters were the width of the gaussian (representing functional precision—see ref. [22], 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.

Resting state analysis. We collected 4 sequential runs of resting state. Each run was 10 min in length, with the subject fixating on a crosshair throughout. Minimal processing steps, performed with AFNI, were applied to the Standard and NORDIC data. These included slice timing correction and motion correction to the first volume of the first run of the Standard data for both Standard and NORDIC data. For both reconstructions, motion was computed (and corrected) relative to the first volume of the first run of the Standard data. Next, we regressed out the 6 estimates of motion parameters and polynomials up to 5th order. A spherical seed, with radius of 3 mm was placed in the medial prefrontal cortex, corresponding to a location within the Default Mode Network. The extracted seed time course for each run was used to generate a map of Pearson's r values, corresponding to the correlation of each voxel in the brain with the seed timeseries (i.e. seed-based correlation).

Denoising algorithms comparison. We compared the performance of NORDIC to that of other denoising strategies on the 0.5 mm isotropic functional data, which, amongst the many datasets in this paper, represents the one most greatly affected by thermal noise and therefore an ideal candidate for NORDIC. Specifically, we evaluated the performances of a global PCA-based algorithm[26] (PCAwn), and a local PCA-based algorithm DWIDenoise (DWIdn)[27], on both magnitude dicoms and complex dicoms. DWIdn is a publicly available implementation on the MPPCA method[27]. PCAwn was implemented following Thomas et al.[26] by first selecting all voxels in the brain by image intensity thresholding on the average series, and then applying a SVD on the Casorati matrix of the whole volume. Following the SVD, each of the left singular basis vectors was evaluated for signal contributions using the multi-taper analysis[25,26], and an empirical threshold, determined from the ratio of the power to the standard deviation of the power spectrum from the multi-taper analysis[25,26], was utilized to select and remove components which only contributed to the thermal noise as in ref. [26]. DWIdn is a local PCA method designed to select and remove components which only contributed to the thermal noise using an objective threshold derived from the Marchenco–Pastur distribution for random matrices. DWIdn was applied both on magnitude only and on complex dicoms using the MRtrix3 toolbox; http://www.mrtrix.org, with its default optimized settings[27]. To obtain complex dicoms, we converted phase images to radians and then combined them with magnitude

images using mrcalc, a tool also part of MRtrix3. Of these methods NORDIC and DWIdn are the most similar, but presenting a number of importance difference, including, but not limited to, the approach of the threshold selection and the optimization of local patch-size (see ref. [40] for more details).

The impact of denoising methods on fMRI data entailed comparing a number of metrics all described in the previous pages. These included the *t*-maps for the contrast target >surround; the distribution of these *t*-values on an ROI hand drawn on the co-registered *T*1, approximately corresponding to the representation of the target region in V1; smoothness metric as implemented in AFNI and the impact of the different methods on single EPI image quality. The reason for hand drawing the ROI rather than deriving it from the maps themselves, was to ensure no bias towards a specific denoising algorithm. The ROI was further constrained to only include values within the brain for the EPI images, to account for potential misregistration across modalities. The results of these comparisons are presented in Supplementary Fig. 16.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The data that support the findings of this study are available from the corresponding author upon reasonable request, subject to human subjects IRB limitations. Source data are provided with this paper.

Code availability
The codes that support the findings of this study will be available here: https://github.com/SteenMoeller/NORDIC_Raw and on Zenodo (https://doi.org/10.5281/zenodo.5147408 https://zenodo.org/record/5147408). Source data are provided with this paper.

References
1. Ogawa, S. et al. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc. Natl Acad. Sci. USA* **89**, 5951–5955 (1992).
2. Kwong, K. K. et al. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl Acad. Sci. USA* **89**, 5675–5679 (1992).
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. *Magn. Reson. Med.* **25**, 390–397 (1992).
4. Van Essen, D. C. & Ugurbil, K. The future of the human connectome. *Neuroimage* **62**, 1299–1310 (2012).
5. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
6. Jorgenson, L. A. et al. The BRAIN Initiative: developing technology to catalyse neuroscience discovery. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, https://doi.org/10.1098/rstb.2014.0164 (2015).
7. Ugurbil, K. Imaging at ultrahigh magnetic fields: History, challenges, and solutions. *Neuroimage* **168**, 7–32 (2018).
8. De Martino, F. et al. The impact of ultra-high field MRI on cognitive and computational neuroimaging. *Neuroimage* **168**, 366–382 (2018).
9. 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. *Neuroimage* **168**, 345–357 (2018).
10. Hoult, D. I. & Richards, R. E. The signal-to-noise ratio of the nuclear magnetic resonance phenomenon. *J. Magn. Reson.* **24**, 71–85 (1976).
11. Edelstein, W. A., Glover, G. H., Hardy, C. J. & Redington, R. W. The intrinsic signal-to-noise ratio in NMR imaging. *Magn. Reson. Med.* **3**, 604–618 (1986).
12. Pruessmann, K. P., Weiger, M., Scheidegger, M. B. & Boesiger, P. SENSE: sensitivity encoding for fast MRI. *Magn. Reson. Med.* **42**, 952–962 (1999).
13. Triantafyllou, C. et al. Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. *Neuroimage* **26**, 243–250 (2005).
14. Triantafyllou, C., Polimeni, J. R. & Wald, L. L. Physiological noise and signal-to-noise ratio in fMRI with multi-channel array coils. *Neuroimage* **55**, 597–606 (2011).
15. Bianciardi, M. et al. Sources of functional magnetic resonance imaging signal fluctuations in the human brain at rest: a 7 T study. *Magn. Reson. Imaging* **27**, 1019–1029 (2009).
16. Kay, K. N., Rokem, A., Winawer, J., Dougherty, R. F. & Wandell, B. A. GLMdenoise: a fast, automated technique for denoising task-based fMRI data. *Front. Neurosci.* **7**, 247 (2013).
17. Caballero-Gaudes, C. & Reynolds, R. C. Methods for cleaning the BOLD fMRI signal. *Neuroimage* **154**, 128–149 (2017).
18. Lund, T. E., Madsen, K. H., Sidaros, K., Luo, W. L. & Nichols, T. E. Non-white noise in fMRI: does modelling have an impact? *Neuroimage* **29**, 54–66 (2006).
19. Murphy, K., Birn, R. M. & Bandettini, P. A. Resting-state fMRI confounds and cleanup. *Neuroimage* **80**, 349–359 (2013).
20. Ugurbil, K. et al. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *Neuroimage* **80**, 80–104 (2013).
21. Barth, M., Breuer, F., Koopmans, P. J., Norris, D. G. & Poser, B. A. Simultaneous multislice (SMS) imaging techniques. *Magn. Reson. Med.* **75**, 63–81 (2016).
22. 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. *Neuroimage* **35**, 539–552 (2007).
23. Wilcox, R. R. *Introduction to Robust Estimation of Hypothesisi Testing* (Elsevier Academic Press, 2005).
24. Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C. & Taylor, P. A. FMRI clustering in AFNI: false-positive rates redux. *Brain Connect.* **7**, 152–171 (2017).
25. Mitra, P. P., Ogawa, S., Hu, X. & Ugurbil, K. The nature of spatiotemporal changes in cerebral hemodynamics as manifested in functional magnetic resonance imaging. *Magn. Reson. Med.* **37**, 511–518 (1997).
26. Thomas, C. G., Harshman, R. A. & Menon, R. S. Noise reduction in BOLD-based fMRI using component analysis. *Neuroimage* **17**, 1521–1537 (2002).
27. Veraart, J. et al. Denoising of diffusion MRI using random matrix theory. *Neuroimage* **142**, 394–406 (2016).
28. Hu, X. & Kim, S. G. Reduction of signal fluctuation in functional MRI using navigator echoes. *Magn. Reson. Med.* **31**, 495–503 (1994).
29. 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. *Neuroimage* **44**, 448–454 (2009).
30. 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. *Nat. Neurosci.* **9**, 23–25 (2006).
31. Smith, S. M. et al. Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl Acad. Sci. USA* **106**, 13040–13045 (2009).
32. Kruger, G. & Glover, G. H. Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magn. Reson. Med.* **46**, 631–637 (2001).
33. Kruger, G., Kastrup, A. & Glover, G. H. Neuroimaging at 1.5 T and 3.0 T: comparison of oxygenation-sensitive magnetic resonance imaging. *Magn. Reson. Med.* **45**, 595–604 (2001).
34. Hyde, J. S., Biswal, B. B. & Jesmanowicz, A. High-resolution fMRI using multislice partial k-space GR-EPI with cubic voxels. *Magn. Reson. Med.* **46**, 114–125 (2001).
35. 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. *Neuroimage* **24**, 738–750 (2005).
36. Bodurka, J., Ye, F., Petridou, N., Murphy, K. & Bandettini, P. A. Mapping the MRI voxel volume in which thermal noise matches physiological noise—implications for fMRI. *Neuroimage* **34**, 542–549 (2007).
37. Gonzalez-Castillo, J., Roopchansingh, V., Bandettini, P. A. & Bodurka, J. Physiological noise effects on the flip angle selection in BOLD fMRI. *Neuroimage* **54**, 2764–2778 (2011).
38. Thomas, C. G. & Menon, R. S. Amplitude response and stimulus presentation frequency response of human primary visual cortex using BOLD EPI at 4 T [In Process Citation]. *Magn. Reson. Med.* **40**, 203–209 (1998).
39. Salimi-Khorshidi, G. et al. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* **90**, 449–468 (2014).
40. Moeller, S. et al. NOise Reduction with DIstribution Corrected (NORDIC) PCA in dMRI with complex-valued parameter-free locally low-rank processing. *Neuroimage* **226**, 117539 https://doi.org/10.1016/j.neuroimage.2020.117539 (2021).
41. Adhikari, B. M. et al. A resting state fMRI analysis pipeline for pooling inference across diverse cohorts: an ENIGMA rs-fMRI protocol. *Brain Imaging Behav.* **13**, 1453–1467 (2019).
42. Adhikari, B. M. et al. Heritability estimates on resting state fMRI data using ENIGMA analysis pipeline. *Pac. Symp. Biocomput.* **23**, 307–318 (2018).
43. Chen, J. E., Polimeni, J. R., Bollmann, S. & Glover, G. H. On the analysis of rapidly sampled fMRI data. *Neuroimage* **188**, 807–820 (2019).
44. Triantafyllou, C., Hoge, R. D. & Wald, L. L. Effect of spatial smoothing on physiological noise in high-resolution fMRI. *Neuroimage* **32**, 551–557 (2006).

45. Chaimow, D., Yacoub, E., Ugurbil, K. & Shmuel, A. Spatial specificity of the functional MRI blood oxygenation response relative to neuronal activity. *Neuroimage* **164**, 32–47 (2018).

46. Ugurbil, K. What is feasible with imaging human brain function and connectivity using functional magnetic resonance imaging. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, https://doi.org/10.1098/rstb.2015.0361 (2016).

47. Menon, R. S. Postacquisition suppression of large-vessel BOLD signals in high-resolution fMRI. *Magn. Reson. Med.* **47**, 1–9 (2002).

48. Jochimsen, T. H., Norris, D. G., Mildner, T. & Moller, H. E. Quantifying the intra- and extravascular contributions to spin-echo fMRI at 3 T. *Magn. Reson. Med.* **52**, 724–732 (2004).

49. Duong, T. Q. et al. Microvascular BOLD contribution at 4 and 7 T in the human brain: gradient-echo and spin-echo fMRI with suppression of blood effects. *Magn. Reson. Med.* **49**, 1019–1027 (2003).

50. Zimmermann, J. et al. Mapping the organization of axis of motion selective features in human area MT using high-field fMRI. *PLoS ONE* **6**, e28716 (2011).

51. BRAIN Initiative Working Group. *BRAIN 2025: A Scientific Vision*, http://braininitiative.nih.gov/pdf/BRAIN2025_508C.pdf (2014).

52. Sadeghi-Tarakameh, A. et al. In vivo human head MRI at 10.5: a radiofrequency safety study and preliminary imaging results. *Magn. Reson. Med.* **84**, 484–496 (2020).

53. Ugurbil, K. et al. Brain imaging with improved acceleration and SNR at 7 Tesla obtained with 64-channel receive array. *Magn. Reson. Med.* **82**, 495–509 (2019).

54. Tavaf, N. et al. A self-decoupled 32-channel receive array for human-brain MRI at 10.5 T. *Magn. Reson. Med.* https://doi.org/10.1002/mrm.28788 (2021).

55. Yaman, B. et al. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magn. Reson. Med.* **84**, 3172–3191 (2020).

56. Akcakaya, M., Moeller, S., Weingartner, S. & Ugurbil, K. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging. *Magn. Reson. Med.* **81**, 439–453 (2019).

57. Moeller, S. et al. Diffusion imaging in the post HCP era. *J. Magn. Reson. Imaging* https://doi.org/10.1002/jmri.27247 (2020).

58. Breuer, F. A. et al. General formulation for quantitative G-factor calculation in GRAPPA reconstructions. *Magn. Reson. Med.* **62**, 739–746 (2009).

59. Sotiropoulos, S. N. et al. Effects of image reconstruction on fiber orientation mapping from multichannel diffusion MRI: reducing the noise floor using SENSE. *Magn. Reson. Med.* **70**, 1682–1689 (2013).

60. Aja-Fernández, S. & Vegas-Sánchez-Ferrero, G. *Statistical Analysis of Noise in MRI* (Springer, 2016).

61. Katkovnik, Vladimir, Alessandro Foi, Karen Egiazarian & Astola, J. From local kernel to nonlocal multiple-model image denoising. *Int. J. Comput. Vision* (2010).

62. Dabov, K., Foi, A., Katkovnik, V. & Egiazarian, K. Image denoising with block-matching and 3D filtering. In *Electronic Imaging'06, Proc. SPIE*, (ed. Dougherty E.R., Astola J.T., Egiazarian K.O., Nasrabadi N.M., Rizvi S.A.), SPIE-IS&T Electronic Imaging, Vol. 6064 (2006).

63. Li, X., Morgan, P. S., Ashburner, J., Smith, J. & Rorden, C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *J. Neurosci. Methods* **264**, 47–56 (2016).

64. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).

65. Smith, S. M. et al. Resting-state fMRI in the Human Connectome Project. *Neuroimage* **80**, 144–168 (2013).

Acknowledgements

Author contributions

L.V., E.Y., K.U. designed the experiments. L.V. and L.D. acquired data. L.D. analyzed the 3 T data; F.d.M. analyzed the auditory data; L.V. carried out all remaining analyses. S.M. and M.A. conceptualized the denoising methodology, S.M. wrote the denoising algorithm and executed it on the acquired data. L.V., K.U., S.M., M.A. wrote the paper. All authors discussed the results, the manuscript, and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information