




RESEARCH ARTICLE

Evaluation of a personalized functional near infra-red optical tomography workflow using maximum entropy on the mean

Zhengchen Cai¹  | Makoto Uji¹  | Ümit Aydın^{1,2} | Giovanni Pellegrino^{3,4}  | Amanda Spilkin¹ | Édouard Delaire¹ | Chifaou Abdallah^{3,4} | Jean-Marc Lina^{5,6} | Christophe Grova^{1,3,4,6}

¹Multimodal Functional Imaging Lab, Department of Physics and PERFORM Centre, Concordia University, Montréal, Québec, Canada

²Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

³Neurology and Neurosurgery Department, Montreal Neurological Institute, McGill University, Montréal, Québec, Canada

⁴Multimodal Functional Imaging Lab, Biomedical Engineering Department, McGill University, Montréal, Québec, Canada

⁵Département de Génie Électrique, École de Technologie Supérieure, Montréal, Québec, Canada

⁶Centre De Recherches En Mathématiques, Montréal, Québec, Canada

Correspondence

Zhengchen Cai, Multimodal Functional Imaging Lab, Department of Physics and PERFORM Centre, Concordia University, Loyola Campus, Office SP 365.23, 7141 Sherbrooke Street West, Montréal, QC, H4B 1R6, Canada.
Email: zhengchen.cai@mail.concordia.ca

Funding information

Canadian Foundation for Innovation; Canadian Institutes of Health Research, Grant/Award Number: CIHR MOP 133619; Fonds de Recherche du Québec - Nature et Technologies; Fonds de Recherche du Québec - Santé; Horizon Postdoctoral Fellowships of Concordia University; McGill/MNI - Fred Andermann EEG and Epilepsy Fellowship; Natural Sciences and Engineering Research Council of Canada; PERFORM Graduate Scholarship in Preventive Health Research; Strauss Canada Foundation; Savoy Foundation Postdoctoral Fellowships

Abstract

In the present study, we proposed and evaluated a workflow of personalized near infra-red optical tomography (NIROT) using functional near-infrared spectroscopy (fNIRS) for spatiotemporal imaging of cortical hemodynamic fluctuations. The proposed workflow from fNIRS data acquisition to local 3D reconstruction consists of: (a) the personalized optimal montage maximizing fNIRS channel sensitivity to a predefined targeted brain region; (b) the optimized fNIRS data acquisition involving installation of optodes and digitalization of their positions using a neuronavigation system; and (c) the 3D local reconstruction using maximum entropy on the mean (MEM) to accurately estimate the location and spatial extent of fNIRS hemodynamic fluctuations along the cortical surface. The workflow was evaluated on finger-tapping fNIRS data acquired from 10 healthy subjects for whom we estimated the reconstructed NIROT spatiotemporal images and compared with functional magnetic resonance imaging (fMRI) results from the same individuals. Using the fMRI activation maps as our reference, we quantitatively compared the performance of two NIROT approaches, the MEM framework and the conventional minimum norm estimation (MNE) method. Quantitative comparisons were performed at both single subject and group-level. Overall, our results suggested that MEM provided better spatial accuracy than MNE, while both methods offered similar temporal accuracy when reconstructing oxygenated (HbO) and deoxygenated hemoglobin (HbR) concentration changes evoked by finger-tapping. Our proposed complete workflow was made available in the brainstorm fNIRS processing plugin—NIRSTORM, thus providing the opportunity for other researchers to further apply it to other tasks and on larger populations.

KEYWORDS

finger tapping, functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy (fNIRS), maximum entropy on the mean (MEM), near infra-red optical tomography (NIROT), personalized optimal montage

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

1 | INTRODUCTION

Functional near infra-red spectroscopy (fNIRS) noninvasively measures fluctuations of both oxygenated and deoxygenated hemoglobin (i.e., HbO and HbR) in the cerebral cortex with high temporal resolution (Jöbsis, 1977; Scholkmann et al., 2014; Yücel et al., 2021). fNIRS raw data measured by source-detector pairs called channels refer to light intensity changes at specific wavelengths (e.g., 685 nm and 830 nm), modulated by local absorption associated with underlying fluctuations of hemoglobin concentrations. Concentration changes in HbO/HbR in each channel are usually estimated via the modified Beer–Lambert Law (mBLL) (Delpy et al., 1988).

Importantly, the channel space analysis assumes homogeneous concentration changes within the underlying region(s) of interest. Such an assumption actually introduces systematic errors when dealing with focal hemodynamic responses (Boas et al., 2001; Strangman, Franceschini, & Boas, 2003). Near Infra-Red Optical Tomography (NIROT) is a powerful alternative to channel space analysis to overcome this issue, as well as to generate hemodynamic images along the cortical surface (Arridge, 1999; Boas et al., 2001). NIROT avoids the homogeneity assumption by reconstructing the light intensity changes measured in the channel space on the underneath cortical area, before converting them into local hemodynamic HbO/HbR fluctuations. NIROT reconstruction actually consists in solving an ill-posed inverse problem that requires specific regularization to obtain a unique solution.

The first step of NIROT involves the definition of a specific fNIRS channel layout, also called the montage, used for data acquisition, which should allow sufficient density and spatial overlap between channels to allow accurate 3D reconstruction (White & Culver, 2010). A high-density montage involving various channel distances and extensive overlapping between channels was initially proposed in White and Culver (2010) and Zeff, White, Dehghani, Schlaggar, and Culver (2007) and successfully considered in several NIROT studies (Eggebrecht et al., 2012; Eggebrecht et al., 2014; Ferradal et al., 2016; Hassanpour, Eggebrecht, Peelle, & Culver, 2017; Fishell, Burns-Yocum, Bergonzi, Eggebrecht, & Culver, 2019). However, fNIRS devices allowing such high-density montages are not widely accessible and are usually custom-made solutions not available commercially. In previous studies (Machado et al., 2018; Machado, Marcotte, Lina, Kobayashi, & Grova, 2014), our group has proposed an approach entitled—*personalized optimal montage*, which maximizes the fNIRS channel layout's sensitivity to a targeted region of interest along the cortex, suggesting fNIRS as a technique mainly used to accurately and locally explore hemodynamic processes, rather than a whole brain imaging technique. While avoiding the need for a large high-density montage, our approach could provide a personalized fNIRS montage maximizing the detection efficiency as well as maintaining a sufficient number of channels and spatial overlap to allow local reconstruction of NIROT images.

The other important step when considering NIROT is to solve the inverse problem to reconstruct HbO/HbR maps along the cortex from the channel space measurements. This inverse problem is ill-posed

such that there are infinite number of solutions. Therefore, regularization is required to estimate a unique solution. The most widely used inverse problem estimator is the so-called minimum norm estimation (MNE), which was first proposed by Hämäläinen and Ilmoniemi (1994) for electroencephalography (EEG) and magnetoencephalography (MEG) source localization and since then applied in several fNIRS reconstruction studies (Dehghani, White, Zeff, Tizzard, & Culver, 2009; Eggebrecht et al., 2012; Eggebrecht et al., 2014; Zeff et al., 2007). MNE consists of applying Tikhonov regularization to minimize the L2-norm when solving the ill-posed inverse problem.

The maximum entropy on the mean (MEM) framework was first proposed by Amblard, Lapalme, and Lina (2004), before being adapted and carefully evaluated in the context of EEG/MEG source localization (Chowdhury, Lina, Kobayashi, & Grova, 2013; Grova et al., 2006). A key property of MEM source imaging is its ability to accurately recover the spatial extent of the generators, as demonstrated in the context of: (a) localizing transient epileptic discharges (Chowdhury et al., 2016; Grova et al., 2016; Heers et al., 2016; Pellegrino et al., 2016; Pellegrino et al., 2020) and oscillations (Avigdor et al., 2020; Pellegrino, Hedrich, et al., 2016); (b) localizing focal sources, such as those evoked by electrical median nerve stimulations (Hedrich, Pellegrino, Kobayashi, Lina, & Grova, 2017); (c) EEG/MEG fusion in the presurgical evaluation of epilepsy (Chowdhury et al., 2018); and (d) MEG resting state connectivity (Aydin et al., 2020). In our previous study (Cai et al., 2021), we adapted the MEM framework to perform fNIRS reconstructions to generate NIROT images and then carefully evaluated MEM performance within a comprehensive and realistic simulation framework. In this study, we opted to combine the above methodology developments as a workflow for conducting NIROT and evaluated its performance using the real data acquired during a motor task.

Here, we introduce a comprehensive NIROT workflow, *the personalized NIROT using MEM*, to accurately reconstruct and assess HbO/HbR fluctuations within targeted brain regions. Our workflow allows: (a) experiment planning and optimal probe design; (b) personalized fNIRS montage installation and digitalization of sensor positions using the neuronavigation device and (c) reconstruction of hemodynamic images using the MEM method. To evaluate the workflow, we considered a finger tapping task applied on a cohort of 10 healthy participants who performed the task twice, once during fMRI and then during personalized fNIRS data acquisition. fMRI Z-maps served as the reference for validation purposes. Evaluations were conducted by visual inspections of the reconstructed NIROT maps and using several quantitative validation metrics such as Area Under the Receiver Operating Characteristic (ROC) curve (AUC), Minimum geodesic distance (Dmin), Spatial Dispersion (SD) and reconstructed HbO/HbR peak times. We also compared MEM reconstructions to the conventional MNE approach. Finally, we assessed the reliability (Yu et al., 2018; Zuo, Xu, & Milham, 2019) of the performance differences between MEM and MNE, using a non-parametric measure of discriminability (Bridgeford, Shen, Wang, & Vogelstein, 2018; Wang, Bridgeford, Wang, Vogelstein, & Caffo, 2020).

2 | MATERIALS AND METHODS

2.1 | Subject cohort and experiment protocol

Ten healthy subjects (24 ± 5 years old, right-handed) participated in this study. The study was approved by the Central Committee of Research Ethics of the Minister of Health and Social Services Research Ethics Board, (CCER), Québec, Canada. All subjects signed a written informed consent before participating in the study.

The study comprised two sessions for each participant who performed the same task during a fMRI scan and a personalized fNIRS scan. Participants performed a finger-tapping task which consisted of tapping the left thumb to the other four digits sequentially, with a pace at around 2 Hz. The duration of the finger tapping block was 10 s, and blocks were interleaved with a resting period lasting between 30 s to 60 s (i.e., one random sampled resting period in this range per block). In the fMRI scan, during the resting period, participants looked at a fixation cross through a mirror placed above the subjects' forehead. They were also asked not to move the head and body while tapping the finger to reduce motion artifacts. The beginning/end of each block was signaled by an auditory cue. The inter-block interval was set so to reduce the influence caused by physiological and systemic fluctuations so to achieve a better estimation of task-evoked brain hemodynamic responses (Aarabi, Osharina, & Wallois, 2017). The finger tapping-rest sequence was repeated 20 times, for around 16 min scan duration.

2.2 | Structural and functional MRI acquisitions

Each participant underwent anatomical and functional MRI with a General Electric Discovery MR750 3T scanner at the PERFORM Centre of Concordia University, Montréal, Canada. The anatomical scans were performed for fNIRS head model, optimal montage, and coregistration between anatomical and fMRI data. In details, data were recorded with the following parameters:

- T1-weighted anatomical images with the 3D BRAVO sequence ($1 \times 1 \times 1 \text{ mm}^3$, 192 axial slices, 256×256 matrix).
- T2-weighted anatomical images were scanned using the 3D Cube T2 sequence ($1 \times 1 \times 1 \text{ mm}^3$ voxels, 168 sagittal slices, 256×256 matrix).
- fMRI images with a gradient echo EPI sequence ($3.7 \times 3.7 \times 3.7 \text{ mm}^3$, 32 axial slices, TE = 25 ms, TR = 2000 ms, 70° flip angle).

2.3 | Anatomical data processing

High-resolution T1- and T2-weighted images were processed using FreeSurfer 6.0 (Fischl et al., 2002) (<https://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferWiki>) and SPM12 (Penny, Friston, Ashburner, Kiebel, & Nichols, 2011) (<https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>) to segment the head into five tissues (i.e., scalp, skull, Cerebrospinal fluid

(CSF), gray matter and white matter). Gray matter and white matter masks were generated from the cortical/subcortical segmentation of FreeSurfer. Scalp and skull masks were segmented using SPM. All segmentation processes used both T1 and T2 weighted images to achieve more accurate estimations of the tissue types by taking advantage of their complementary contrasts. Anatomical surfaces such as pial surface, gray/white matter interface and mid surface (i.e., a middle layer of the gray matter) were estimated using FreeSurfer (Fischl & Dale, 2000).

2.4 | fMRI data processing

fMRI data were processed using FSL v6.0.0 (<https://fsl.fmrib.ox.ac.uk/fsl/>) (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012). Data preprocessing included the following steps: (a) slice timing correction using interleaved Hanning-windowed interpolation; (b) brain extraction using BET2; (c) head motion correction applying rigid-body transformations (MCFLIRT); (d) spatial smoothing (5 mm FWHM Gaussian kernel); (e) high-pass temporal filtering (45 s cut-off) and (f) registration to the individual T1 anatomical image (FLIRT, linear transformation with 6 degrees of freedom), and normalization (linear affine transformation with 12 degrees of freedom) to the MNI (Montreal Neurological Institute and Hospital) 1 mm standard brain template (i.e., ICBM152).

The first-level general linear model (GLM) analyses were performed using FEAT v6.0 (Woolrich, Ripley, Brady, & Smith, 2001), employing boxcar (10 s) finger-tapping events convolved with the double-gamma Hemodynamic Response Function (HRF). Time series analysis was carried out using FILM (Woolrich et al., 2001) with local autocorrelation correction. The resulting first-level task-evoked BOLD Z statistic images were thresholded using Gaussian random field theory based maximum height thresholding (Worsley, 2001) with the voxel-wise inference ($p < .01$, corrected). To allow visual and quantitative comparison with NIROT, individual images estimated along the cortical surface, individual volumetric fMRI Z-maps were projected onto the mid-surface (downsampled to 25,000 vertices) using a Voronoi-based projection, which is a volume to surface interpolation method able to preserve sulco-gyral morphology (Grova et al., 2006). Z-values were averaged within each Voronoi cell and assigned to the corresponding vertex of the cortical surface.

By using each subject's volume-based fMRI first-level results, we conducted a group-level analysis in order to identify a reliable fMRI reference region of interest (ROI) to be considered as our "ground truth" for NIROT images evaluation. This group-level analysis was performed using a mixed-effects model—FLAME1 in FEAT (Woolrich, Behrens, Beckmann, Jenkinson, & Smith, 2004). The resulted BOLD Z-statistic images were thresholded using Gaussian random field theory based maximum height thresholding (Worsley, 2001) (clusters determined $Z > 3.1$, cluster significance threshold of $p < .01$, corrected). The thresholded fMRI group-level Z-map registered on the MNI standard template (see Figure S1) was projected onto the mid-surface of the template per se, using the same Voronoi projection method previously mentioned. The significant region on the surface was determined according to the Z threshold of volume-based results.

2.5 | Personalized fNIRS data acquisition and preprocessing

fNIRS data acquisition was conducted in the Multimodal Functional Imaging Laboratory at PERFORM Center (Concordia University, Montréal, Canada) using a continuous wave Brainsight fNIRS device (Rogue-Research Inc, Montréal, Canada). Personalized fNIRS acquisition strategy consists of estimating a subject-specific optimal fNIRS montage to maximize, a priori, the fNIRS sensitivity to the hemodynamic activity in some targeted brain regions, while ensuring sufficient spatial coverage and overlap to allow accurate local 3D reconstruction (Machado et al., 2014; Machado et al., 2018). In this study, the hand-knob region within the right primary motor cortex was defined as the individual ROI. It was selected manually along the cortical surface (see Figure 1a), taking into account anatomical landmarks (Raffin, Pellegrino, Di Lazzaro, Thielscher, & Siebner, 2015) and using the Brainstorm software (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011) (<http://neuroimage.usc.edu/brainstorm>). Then, we estimated a personalized optimal fNIRS montage under the following constraints (see Figure 1b): (a) 3 sources and 15 detectors; (b) source-detector distance between 2.0cm and 4.5cm; (c) spatial overlap between channels—signal from each source to be detected by at least 13 out of 15 detectors. For each candidate source/detector pair, a fNIRS forward model, computed using MCXLab Monte Carlo Photon simulator (Fang & Boas, 2009; Yu, Nina-Paravecino, Kaeli, & Fang, 2018) was considered to measure light sensitivity within the target region. The output of the resulting optimization algorithm under proposed constraints, consisted in a set of spatial 3D coordinates indicating the optimal position of the optodes on subject's scalp. Please find further details in (Cai et al., 2021; Machado et al., 2014; Machado et al., 2018). To monitor systemic fluctuations within superficial layers, we also added one proximity detector (i.e., 0.7cm to each source), located in the center of the 3 sources of the montage. These optimal 3D coordinates were loaded into a neuronavigation system (Brainsight TMS navigation system, Rogue-Research Inc, Montreal) to guide the installation of the optodes on the scalp (Machado et al., 2018). fNIRS sensors were then glued on the scalp (see Figure 1d) using a clinical

adhesive, called collodion, in order to minimize sensitivity to motion artifacts and to improve fNIRS signal to noise ratio (Machado et al., 2018; Yücel, Selb, Boas, Cash, & Cooper, 2014).

Raw fNIRS data were then preprocessed using a conventional pipeline. Channels exhibiting either negative raw amplitude or a coefficient of variation (standard deviation over the signal mean) larger than 8% were classified as low-quality and rejected (Eggebrecht et al., 2012; Piper et al., 2014; Schmitz et al., 2005; Schneider et al., 2011). Superficial physiological fluctuations, also considered here as noise, were modeled by the average signal of all proximity channels and were regressed out from all channels using a linear regression (Zeff et al., 2007). All channels were band-pass filtered between 0.01Hz and 0.1Hz using a third order Butterworth filter. Optical density changes (i.e., ΔOD), normalized for each channel by the mean amplitude of the entire time course, were calculated using the logarithm conversion of the filtered signal. ΔOD epochs with a time window ranging from $-10s$ to $30s$ around the task onset were extracted, and noisy epochs were rejected through visual inspections (i.e., overall, less than 4 out of 20 epochs were rejected for each subject, mainly caused by motion artifacts). Then, the resulting epochs were averaged. This epoch averaged ΔOD , measured either at 685 nm or 830 nm, was the input signal of fNIRS reconstruction algorithm to obtain the NIROT images for each subject.

2.6 | Forward model estimation for NIROT

The subject-specific head model was calculated to estimate how a local change in light absorption elicited by the hemodynamic response at a specific vertex of the cortical area would impact fNIRS signals on scalp channels. Volume-based subject-specific head segmentation (e.g., scalp, skull, CSF, gray matter and white matter), were used to model the light propagation from each optode of the optimal montage within head tissues. We assigned each tissue the same optical coefficients used in Fang & Boas, 2009; Machado et al., 2018 for the two wavelengths (i.e., 685 nm and 830 nm). Fluences of each wavelength for each optode were calculated by simulating the propagation of 10^8 photons

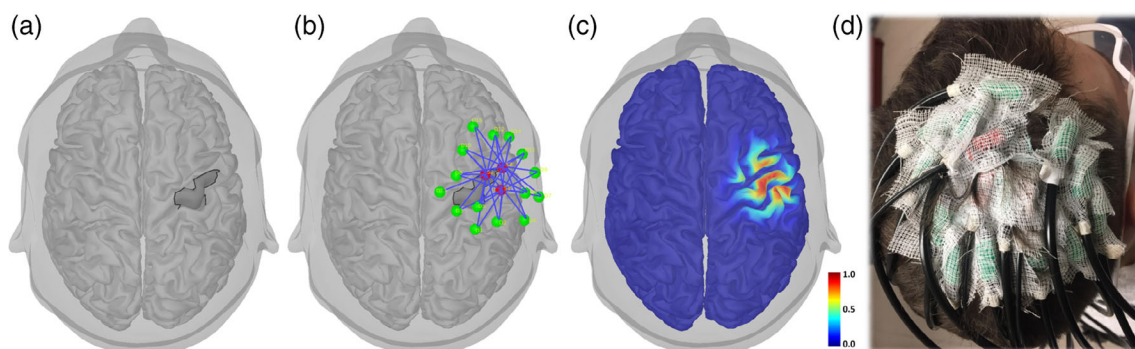


FIGURE 1 Personalized fNIRS investigation using an optimal montage targeting the right motor cortex. (a) Black profile represents the right “hand knob” targeted region of interest, (b) resulting personalized optimal montage targeting this ROI consisting of 3 sources (red), 15 detectors (green) and one proximity (in the center of sources not shown), (c) normalized light sensitivity profile of the optimal montage which was calculated as the sum of all channel's sensitivity along the cortical surface, (d) optimal montage glued on the subject's head using collodion. Installation was guided by a neuronavigation system, and optodes positions were then digitized after the acquisition for coregistration and forward modeling

using MCXLAB (Fang & Boas, 2009; Yu, Nina-Paravecino, et al., 2018) (<http://mcx.space/wiki/index.cgi?MCX>). Each voxel's sensitivity value corresponding to each fNIRS channel was then computed using the adjoint formulation with the Rytov approximation (Arridge, 1999).

To constrain the fNIRS reconstruction space within the cortical region (Boas & Dale, 2005), we applied volume-to-surface interpolation using Voronoi (Grova, Makni, et al., 2006) to generate surface based sensitivity matrix along the mid cortical surface (please see Cai et al., 2021 for further details).

Finally, to define the field of view (FOV) of fNIRS reconstruction, this surface-based sensitivity was further spatially constrained to be within 3 cm to any optode of a specific optimal montage, Euclidean distance calculated from 100% inflated cortical surface to the head surface by Brainstorm (Tadel et al., 2011).

2.7 | NIROT reconstruction

We compared two fNIRS reconstruction methods, one being our previously proposed and validated maximum entropy on the mean (MEM) (Cai et al., 2021) and the other one being the conventional minimum norm estimation (MNE) (Hämäläinen & Ilmoniemi, 1994). MEM offers an efficient nonlinear probabilistic Bayesian framework to incorporate prior knowledge in the solution of the inverse problem. It assumes that brain activity is modeled by cortical parcels that are estimated using a data-driven parcellation of the full field of view considered for NIROT. While fitting the data through relative entropy maximization, MEM has the unique ability to switch off parcels of the model considered as inactive using a hidden variable. In our previous studies in the context of EEG/MEG source imaging, we have demonstrated excellent accuracy of MEM and the ability to be sensitive to the spatial extent of the underlying generators (Chowdhury et al., 2013; Chowdhury et al., 2016; Grova et al., 2016; Hedrich et al., 2017; Heers et al., 2016; Pellegrino et al., 2020), before adapting this framework in the context of NIROT (Cai et al., 2021). As the most conventional inverse procedure considered in NIROT, MNE is a linear method (Hämäläinen & Ilmoniemi, 1994) using Tikhonov regularization to minimize the L2-norm. Please refer to Appendix A for further details on those NIROT methods. In practice, NIROT reconstructions were calculated by our implementations of these two methods in the fNIRS processing plugin—NIRSTORM (<https://github.com/Nirstorm/nirstorm>) in Brainstorm software.

2.8 | Quantitative evaluation by comparing NIROT to fMRI

To evaluate and compare the spatial accuracy of NIROT reconstructions obtained using either MEM or MNE mentioned above, we proposed to compare the reconstructed HbO/HbR responses along the cortical surface to surface-based fMRI Z-maps, both at the individual and at the group level. To do so, we first defined two cortical regions

from the resulted surface-based fMRI Z-map as the ROIs to quantitatively assess the spatial accuracy of NIROT maps.

2.8.1 | Determination of the cortical ROIs used for quantitative evaluation

ROIs selected for quantitative evaluation consisted of two cortical regions, extracted from surface-based subject-specific fMRI Z-map response to the finger-tapping task. The first one, fMRI reference ROI#1, corresponded to the “activated” region where the main activation of task-evoked HbO/HbR responses should be expected. The second one, fMRI reference ROI#2, corresponded to the “non-activated” region where no significant HbO/HbR responses were expected. These two fMRI regions were our reference “true positives” and “true negatives” for later NIROT evaluations.

The determination of fMRI reference ROI#1 for each subject was obtained from the first-level Z-map projected on individual mid-surface. We did not directly use the significant region on the projected Z-map (threshold) considering that each subject only went through one fMRI task session, and this might not be enough to estimate a reliable size of the main activation cluster, as suggested by fMRI test-retest reliability studies during motor tasks (Quiton, Keaser, Zhuo, Gullapalli, & Greenspan, 2014; Zandbelt et al., 2008). Instead, the ROI#1 was defined by thresholding the individual surface-based fMRI Z-map until a predefined size, in order to ensure a similar spatial extent of the activated area of the one found at the group level. The group level analysis provided a more reliable generalized size parameter of the activation pattern, as suggested by Zandbelt et al. (2008). Therefore, to identify the expected size of fMRI reference ROI#1, we considered from the fMRI group-level map (described in Section 2.4) the most significant cluster (Figure 2a) projected along the cortical surface. It resulted in 366 vertices shown by the black profile in Figure 2a. Consequently, we considered a parameter size of 366 vertices to determine the threshold of each individual map, to estimate the “activated” region at the expected size. It is worth mentioning that such an approach was actually more conservative than the conventional threshold obtained at the single subject volume map (see Section 2.4, $p < .01$ corrected using Gaussian random field theory). Finally, fMRI reference ROI#2 (“true negatives”) was directly defined as the “non-activated” region (see Figure 2b) on each individual fMRI Z-map on the cortical surface, by identifying vertices that were below the conventional threshold on the single subject volume map ($p < .01$ corrected using Gaussian random field theory). The remaining vertices between ROI#1 and ROI#2 could be considered as a region of fMRI uncertainty at the single subject level.

2.8.2 | Quantitative validation metrics

To assess the spatial accuracy of NIROT maps, we applied similar validation metrics described in our previous evaluation of MEM source imaging in the context of EEG/MEG data (Chowdhury et al., 2013;

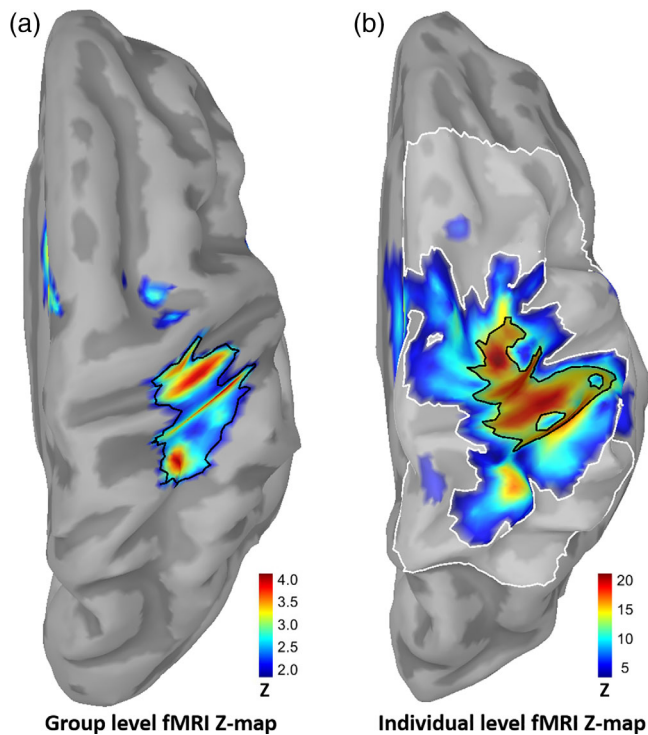


FIGURE 2 Determination of the cortical regions used for quantitative evaluations. (a) group-level fMRI Z-map obtained by FSL group-level activation map projected along the mid-surface of MNI template ICBM152. The map was thresholded to only contain the significant activation region along the surface. Black profile represents the most significant cluster, which consists of 366 vertices. (b) Determination of fMRI reference ROIs on the individual level surface-based fMRI Z-map of Sub01, projected from FSL first-level activation map on subject's mid-surface. The black profile is the individual fMRI reference ROI#1 which contains 366 vertices when thresholding the individual map by Z values. It represents the “activated” region used for AUC, SD and Dmin metrics. The white profile is the individual fMRI reference ROI#2 represents the intersection between nonsignificant region and NIROT reconstruction field of view, which is defined as the “non-activated” region for AUC calculations. Cortical surfaces are 50% inflated for visualization purposes

Chowdhury et al., 2016; Grova, Makni, et al., 2006; Hedrich et al., 2017). Each reconstructed NIROT image is actually a spatiotemporal map of the whole hemodynamic response, whereas the fMRI Z-map consists of a statistical static map assessing the presence of a modeled canonical hemodynamic response. To select a single NIROT map to evaluate, from the spatiotemporal estimates of each subject using each reconstruction method, we first averaged the reconstructed HbO/HbR time courses within the fMRI reference ROI#1, and then extracted the HbO/HbR maps at the time point of HbO versus HbR peaks on these averaged time courses. Moreover, since we have no ground truth for the reconstructed time course in fNIRS and since fNIRS is considered a relative measurement compared to baseline activity, we considered only the shape of the reconstructed fNIRS response by normalizing each selected spatial-temporal HbO/HbR map by this selected peak amplitude.

Using the 2 previous selected fMRI reference ROIs, defining our reference “true positives” and “true negatives,” we proposed the following four validation metrics to evaluate NIROT spatial and temporal accuracy, for both MNE and MEM:

- **Area Under the Receiver Operating Characteristic (ROC) curve (AUC)** was used to assess the detection ability of the reconstruction methods. ROC analysis consists of assessing the sensitivity and specificity of NIROT maps, when varying the threshold from 0 to maximum amplitude of normalized NIROT maps at their peak. The ROC curve is obtained by plotting sensitivity as a function of $(1 - \text{specificity})$. AUC score is then estimated as the area under this ROC curve. It is considered as a measure of detection accuracy, assessing the sensitivity of NIROT map to the underlying spatial extent of the hemodynamic response. We used the two ROIs defined in the previous section to calculate sensitivity and specificity. Sensitivity was calculated when considering fMRI reference ROI#1 as the “true positive (TP)” region, and specificity was calculated when considering fMRI reference ROI#2 as the “true negative (TN)” region. Moreover, we applied a random and homogeneous parcellation on ROI#2 to have the size of each cluster similar to the size of ROI#1. Then the fictive generators were sampled cluster wise within ROI#2, instead of vertex wise when calculating the True Negative rate. This approach was considering the fact that if the fictive generators would have sampled vertex wise, the number of samples for calculating True Negatives would have been much larger than for calculating True Positive. This procedure is consistent with Grova, Daunizeau, et al. (2006) on AUC calculation to fit the context of a distributed source model—generating as many “fictive” sources in the TN region as the number of vertices of the TP region. Therefore, our approach to estimate AUC prevented underestimation of the True Negative rate. Please note that the remaining vertices between selected ROI#1 and ROI#2 were not considered by our proposed ROC analysis, but they were actually taken into account in the following Dmin and SD metrics.
- **Minimum geodesic distance (Dmin)** was calculated as the minimum geodesic distance, following the circumvolutions of the cortical surface, from the vertex that exhibited maximum of reconstructed activity to the closest border from the individual fMRI reference ROI#1. The Dmin score would therefore be 0 mm if the peak of the HbO/HbR reconstruction map was located inside the fMRI reference ROI#1.
- **Spatial dispersion (SD)** assessed the spatial spread of the estimated activation region distribution and the localization error using the

$$\text{following equation: } SD = \sqrt{\frac{\sum_{i=1}^N (\min_{j \in \Theta} (D^2(i,j))) \hat{X}^2(i,\tau)}{\sum_{i=1}^N (\hat{X}^2(i,\tau))}}$$

where $\min_{j \in \Theta} (D^2(i,j))$ is the minimum squared Euclidean distance between the vertex i to a vertex j located inside the fMRI reference ROI#1 represented by Θ . $\hat{X}^2(i,\tau)$ is the power of the peak amplitude

of the reconstructed time course on vertex i at the peak time τ . N is the total number of vertices within the reconstruction field of view. The ideal value (i.e., $SD = 0mm$) would be achieved when no activation is reconstructed outside the fMRI reference ROI#1. The larger the SD is, the more spatially spreading or mislocalized the reconstructed maps are.

- **Peak times** of reconstructed HbO/HbR time course and the delay between HbO and HbR peak times were compared when reconstructing NIROT images using MEM and MNE. The average of the reconstructed time course of HbO/HbR within fMRI reference ROI#1 was first calculated to extract corresponding peak times (i.e., $Peak_{HbO}$ and $Peak_{HbR}$). The delay between HbO and HbR peak times was then calculated as $Peak_{HbR} - Peak_{HbO}$.

To conduct group-level evaluation for NIROT, individual NIROT maps at their corresponding peak amplitude were first normalized to $[-1, 1]$. FreeSurfer spherical registration was used to project normalized individual HbO/HbR maps onto the mid surface of ICBM152 template. Group-level HbO/HbR reconstruction maps were then estimated by averaging the projected individual maps. AUC, SD and Dmin were calculated by comparing to the group-level fMRI Z-map. We also conducted in the supplementary materials a group-level Z map, in which the Z score was calculated for each vertex, using the mean amplitude divided by the standard deviation among 10 subjects, instead of considering only the average.

Additionally, AUC was also calculated under three different scenarios to answer the following questions specific to the use of personalized fNIRS reconstructions.

- *Was MEM more accurate than MNE at different time instants other than only the peak of the hemodynamic response?* AUCs of the reconstructed HbO/HbR maps using MEM and MNE were evaluated at different time instants ranging from 5 s to 20 s with a step size of 0.5 s.
- *Would it be possible to further threshold MNE reconstruction maps to achieve a similar spatial accuracy than the one obtained when using MEM?* AUCs of HbO/HbR peak maps were calculated by considering different initial thresholds, as a percentage of the peak amplitude of each corresponding map, ranging from 0% to 50% with an increment of 1%, instead of standard AUC starting with a threshold of 0%.
- *Was the 3D reconstruction obtained by using the combination of personalized optimal montage and MEM reconstruction primarily biased by the local forward model (i.e., sensitivity map), since the optimal montage was optimized to a targeted ROI in the motor region?* We assumed the sensitivity profile of each subject, obtained by summing up the sensitivity of all channels of the corresponding optimal montage, to mimic a fully “montage-biased” reconstruction map. Therefore, computing the AUC score of this sensitivity profile and comparing it to the reconstruction results obtained from either MEM or MNE can evaluate whether the local forward model plays a role in biasing the reconstruction.

2.9 | Reliability of the performance differences between MEM and MNE

Considering reliability is the basis for individual difference research, especially on personalized methods (Yu, Linn, et al., 2018; Zuo et al., 2019), we also evaluated the reliability of reconstruction performance of MEM and MNE individually, and more importantly, when considering the paired performance differences between the two methods. To do so, we considered the discriminability measure, which was proposed as a novel nonparametric approach for assessing reliability, proved to be more robust than traditional intraclass correlations and fingerprinting approaches (Wang et al., 2020; Yu, Linn, et al., 2018). Considering a reliability study design, pair-wised distances of the measurement of interest are first calculated for all experiment sessions (within- and between-subjects), and discriminability is then defined as the proportion of the number of cases in which within-subject distances are smaller than between-subject distances. We applied the Multiscale Graph Correlation (MGC) package (Bridgeford et al., 2018; Vogelstein et al., 2019) (available at <https://github.com/neurodata/r-mgc>) to estimate discriminability. Euclidean distance was considered as the default distance measurement to construct within- vs. between-subject distance matrices (Wang et al., 2020) (see Figure B2).

Although we did not have multiple sessions of the same task for each subject, we considered random sub-averaged 16 nonoverlapping blocks out of a total of 20 blocks to mimic the task performance variability within each subject. To ensure good coverage of within-subject variance, we sorted all sub-averaged trials by decreasing SNR values and selected 10 trials with a step of 10 SNR value increments around the median SNR value. The selection of 10 sub-averaged trials was to ensure the same degree of freedom of within-subject when compared to between-subject variance, which was 10 subjects in our case. The selection of 16 blocks was to ensure not reducing SNR too much after averaging. This number was empirically defined according to the observation that usually there were less than 4 artifacts contaminated blocks in one finger tapping session. Selecting sub-averaged trials around the median SNR ensured a good representation of fNIRS responses, while discarding artifacts in the meantime. For instance, in artifacts contaminated data, large motion artifacts would result in high SNR of sub-averaged trials. We then performed both MEM and MNE reconstructions on all data samples (e.g., $100 = 10$ within-session sub-averaged $\times 10$ subjects) and applied the same evaluations for the reconstructed hemodynamic maps. Discriminability was then estimated for each validation metric AUC, SD and Dmin, considering either MEM and MNE individually, or paired performance differences between them.

3 | RESULTS

3.1 | Reconstruction performance comparisons at the individual-level

Statistical summary of individual level results is presented in Figure 3. Paired two sample t tests were considered when comparing

performance between the two methods (MEM vs. MNE). When reconstructing HbO responses, MEM provided significantly larger AUC values than MNE (0.79 vs. 0.68, $p < .01$). Moreover, MEM also showed significantly lower spatial dispersion (SD) than MNE (5.11 mm vs. 9.83 mm, $p < .01$). Both methods reconstructed the peak amplitude very close from the fMRI main cluster as quantified using Dmin, with no significant difference between MEM and MNE (2.17 mm vs. 3.00 mm, $p > .05$). Similar performances were also observed when reconstructing HbR responses. MEM provided significantly larger AUC (0.80 vs. 0.70, $p < .05$), significantly lower SD (5.93 mm vs. 8.97 mm, $p < .05$) and similar Dmin values (3.33 mm vs. 2.90 mm, $p > .05$).

Figures 4 and 5 present all the individual level reconstruction maps and time courses calculated using MEM and MNE for each of 10 subjects. Figures 4a and 5a showed the individual fMRI Z-map along with the individual fMRI reference ROI#1 represented by a black profile. All subjects showed clear fMRI activations along M1 and S1 areas, evoked by the finger tapping task. Columns b and c of Figures 4 and 5 showed NIROT maps normalized by the peak amplitude of HbO/HbR reconstructed using MEM and MNE, respectively, together with validation metric values. For Sub02, 03, 05, 07, 08, 09 and 10, MEM reconstructed accurately HbO/HbR responses with high AUC values ($[Q_1, Q_3] = [0.73, 0.89]$ for HbO, $[0.74, 0.94]$ for HbR among all maps, Q_1, Q_3 referring to the first and third quartile of the distributions, respectively) and small spatial spread (low SD values, $[Q_1, Q_3] = [3.52 \text{ mm}, 6.63 \text{ mm}]$ for HbO, $[3.45 \text{ mm}, 6.37 \text{ mm}]$ for HbR). On the other hand, for these selected 7 subjects, MNE provided less accurate reconstructions characterized by lower AUC values ($[Q_1, Q_3] = [0.60, 0.75]$ for HbO, $[0.60, 0.77]$ for HbR) and larger SD values ($[Q_1, Q_3] = [7.64 \text{ mm}, 10.46 \text{ mm}]$ for HbO, $[8.31 \text{ mm}, 10.59 \text{ mm}]$ for HbR). For Sub01, MEM and MNE were both able to provide accurate HbO/HbR reconstructions when comparing to fMRI results. For Sub06, MEM and MNE provided similar reconstruction results, all maps were able to recover the main cluster but spread out toward more anterior regions. For Sub04, only MEM provided a good

reconstruction only for HbO, resulting in an AUC = 0.7 and Dmin = 0 mm. Column d of Figures 4 and 5 illustrated the averaged reconstructed HbO/HbR time courses within the fMRI reference ROI#1. Temporal fluctuations of averaged reconstruction time courses of HbO/HbR were similar between MEM and MNE in most of the subjects, except that the amplitudes were larger for MEM in Sub 04, 07 and 08, because MNE did not reconstruct the spatial map accurately within the expected region.

Table 1 summarized the statistical comparison of the peak times of HbO/HbR extracted from the above averaged reconstruction time courses within the fMRI reference ROI#1. When comparing MEM and MNE results, there were no significant differences in the peak times of HbO/HbR (paired two sample t test, $p = .71$ and $.17$ for HbO and HbR, respectively). Both NIROT methods demonstrated a significant (one sample t test against 0, < 0.05) delay between the peak time of HbR and HbO ($0.9 \text{ s} \pm 0.3 \text{ s}$ and $1.9 \text{ s} \pm 0.7 \text{ s}$ for MEM and MNE, respectively). Moreover, there was no significant difference (paired two sample t test, $p = 0.27$) between the delays estimated by MEM and MNE. Note that Sub04 was rejected from this analysis since it was exhibiting a very noisy reconstructed time course and therefore was considered as an outlier (Figure 4d).

3.2 | Reconstruction performance comparisons at the group-level

Figure 6 reports fMRI versus NIROT comparisons at the group-level. Validation metrics were calculated on the group averaged NIROT HbO and HbR maps, when using either MEM or MNE. Similar trends were found when considering group-level comparisons by visualization. When compared to MNE results (Figure 6, using fMRI group-level activation as a reference), MEM provided similar AUC (0.73 vs. 0.72 for HbO and 0.74 vs. 0.74 for HbR), lower SD (7.52 mm vs. 10.52 mm for HbO and 7.48 mm vs. 11.18 mm for HbR) and similar Dmin (6.14 mm vs. 6.14 mm for HbO and 2.99 mm vs. 2.76 mm

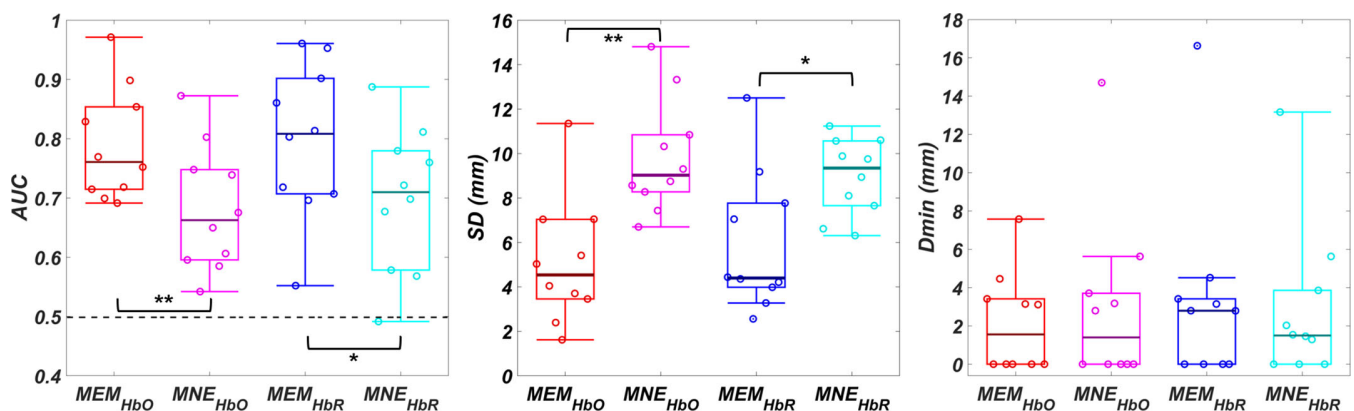


FIGURE 3 Quantitative evaluation of the performances of MEM and MNE reconstructions. (a) Distribution of AUC values for MEM and MNE NIROT reconstructions, for HbO and HbR responses respectively, (b) distribution of SD (in mm) comparison between MEM and MNE in the context of reconstructed HbO/HbR, (c) distribution of Dmin (in mm) comparison between MEM and MNE in the context of reconstructed HbO/HbR. Each circle represents the index of one subject, superimposed on a boxplot representation of the distribution, * represents significant paired two sample t test at $p < .05$ and ** for $p < .01$

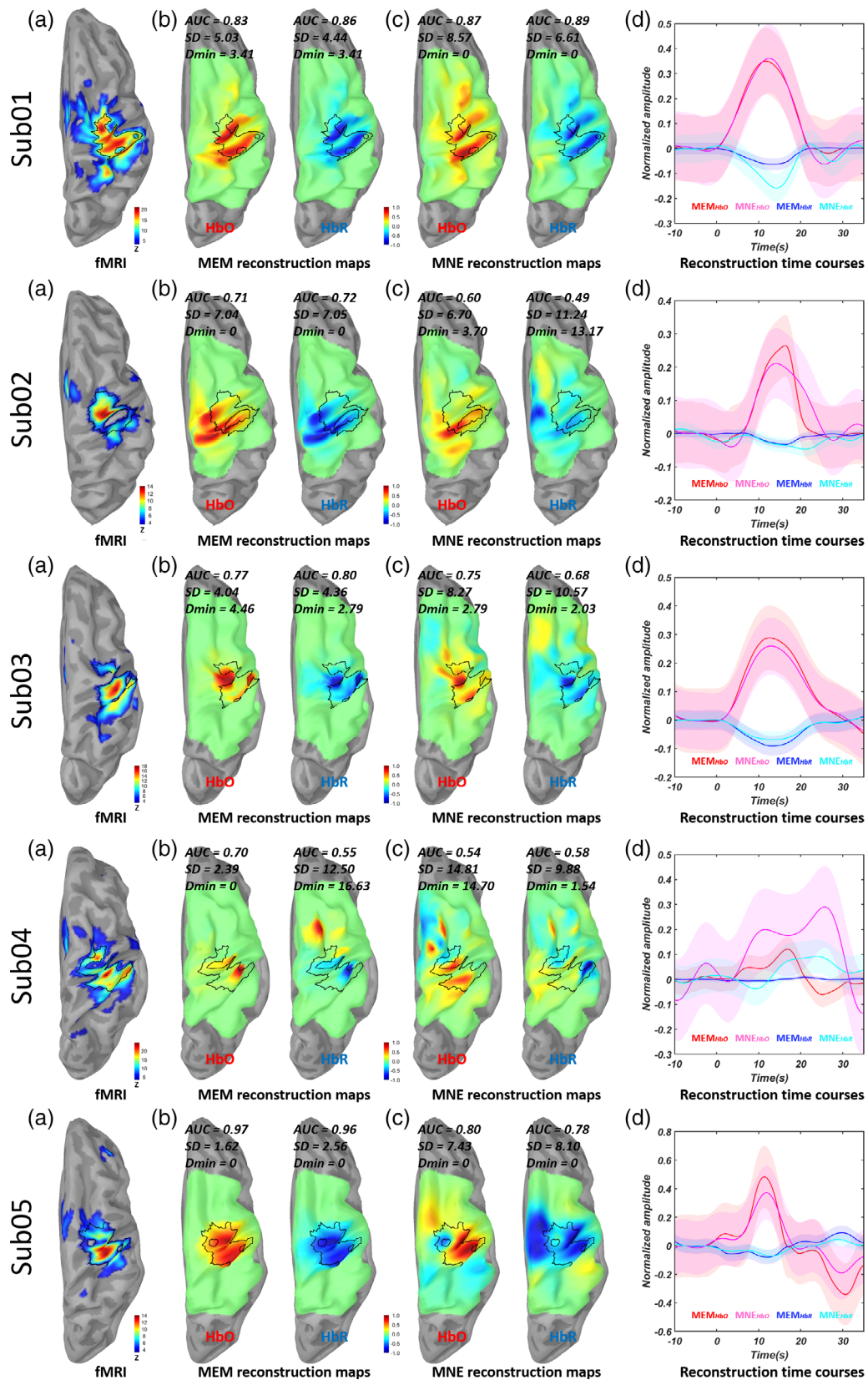


FIGURE 4 Legend on next page.

for HbR). When comparing the group averaged reconstructed HbO/HbR time course within the fMRI main cluster region, as a nonlinear method MEM provided almost identical ones to MNE. As a reference, we also illustrated the expected fMRI time course (black dash line in Figure 6d) by convolving a standard canonical hemodynamic response function (HRF) (Penny et al., 2011) and a 10s boxcar stimuli. It followed well both MEM and MNE reconstructed HbR mean time courses and laid within the shaded areas, demonstrating the good temporal accuracy of our NIROT workflow.

3.3 | NIROT performance along time

Figure 7a illustrates the comparison of AUC values obtained for MEM and MNE reconstructions when considering NIROT maps at different time samples between 5 s and 20 s by steps of 0.5 s. Overall, AUC values corresponding to HbO and HbR reconstructions, when considering either MEM or MNE, were following the temporal fluctuation of the expected task-evoked hemodynamic responses from 5 s to 20 s. Within this temporal window, MEM always provided larger AUC values when compared to MNE, for both HbO and HbR reconstructions. The differences were statistically significant (paired two sample *t* test at each time sample, $p < .05$ for both HbO and HbR comparisons) within the range 8.5 s to 14.5 s. Additionally, AUC values of HbR were also slightly larger than HbO for both reconstruction methods.

3.4 | NIROT performance when applying spatial thresholding on reconstruction maps

Figure 7b illustrates AUC as a function of different initial percentage thresholds (percentage relative to the peak amplitude of each NIROT map) applied on reconstructed HbO/HbR maps. ROC analyses were estimated from a specific starting percentage amplitude threshold from 0 up to 50%, in order to assess the impact of initial thresholding on detection accuracy. As expected, AUC values decreased when thresholding the reconstruction map with a larger initial percentage threshold. Overall, MEM provided larger AUC than MNE under all the thresholding scenarios, and the difference was statistically significant (paired two sample *t* test at each percentage threshold, $p < .05$ for both HbO and HbR comparisons) within the whole threshold range. Results of this analysis, as illustrated in Figure 7b, showed that

additional thresholding of the MNE results does not improve the estimation accuracy of the activity map, when compared to MEM.

3.5 | AUC of the sensitivity profile of the optimal montage

The last column of Figure 8 demonstrates the AUC values calculated when assessing detection accuracy only on the light sensitivity profile (SP) resulting from the personalized optimal montage of each subject. The mean AUC value among 10 subjects was 0.64, significantly lower than the mean AUC value of MEM reconstructed HbO (0.76, $p < .01$, paired two sample *t* test, Bonferroni corrected) and HbR (0.81, $p < .01$, paired two sample *t* test, Bonferroni corrected). This AUC measure was also lower than MNE reconstructed HbO (0.66) and HbR (0.71) map but the difference was not statistically significant. These results suggest that our detection accuracy is slightly influenced by the installation of the optimal montage, but MEM reconstruction still significantly improves NIROT accuracy within the targeted brain region.

3.6 | Reliability of performance differences between MEM and MNE

Over 100 reconstructions, including 10 within-subject resampled sessions and 10 subjects, all three validation metrics (AUC, SD and Dmin) showed significantly better performances using MEM, when compared to MNE. The 95% confidence interval (CI) of the paired differences (MEM-MNE) was [0.13, 0.17] for AUC, [−5.50, −4.31] for SD (in mm) and [−2.76, −0.23] for Dmin (in mm) (see distributions reported in Figure B1). Discriminability of these performance differences was estimated as 0.68 for AUC, 0.70 for SD and 0.68 for Dmin (see Figure B2), indicating that MEM was exhibiting reliably better performances than MNE. When considering the performance of MEM and MNE individually, the discriminability of MEM was estimated as 0.71 for AUC, 0.66 for SD and 0.64 for Dmin, whereas the discriminability of MNE was estimated as 0.85 for AUC, 0.88 for SD and 0.73 for Dmin.

4 | DISCUSSION

The present study proposed and evaluated a workflow of personalized NIROT using MEM for spatiotemporal imaging of cortical

FIGURE 4 Visual and quantitative evaluation for NIROT images at the individual level (subjects 1 to 5). (a) Individual level fMRI Z-map thresholded to only contain the significant activation of each subject during finger tapping task, black profile represents the fMRI reference ROI#1 (“true positive”), non thresholded gray area (nonsignificant) within the reconstruction FOV represents the fMRI reference ROI#2 (“true negatives”). (b) MEM reconstructed NIROT maps at HbO/HbR peak amplitude, respectively, (c) MNE reconstructed NIROT maps at HbO/HbR peak amplitude, respectively, (d) reconstructed time courses within the black profile, solid lines represent the averaged time courses, and the shaded areas represent standard deviation within the ROI#1. Quantitative evaluation metric results are showed on top of each map, respectively. AUCs were calculated by setting black profiles (fMRI reference ROI#1) as the “activated” region and gray area (fMRI reference ROI#2) as the “non-activated” region. SD and Dmin only considered fMRI ROI#1 as the “activated” region. Note that each subject’s map was normalized by its own peak, and time courses were normalized by the maximum HbO amplitude across two methods before averaging within the black profile

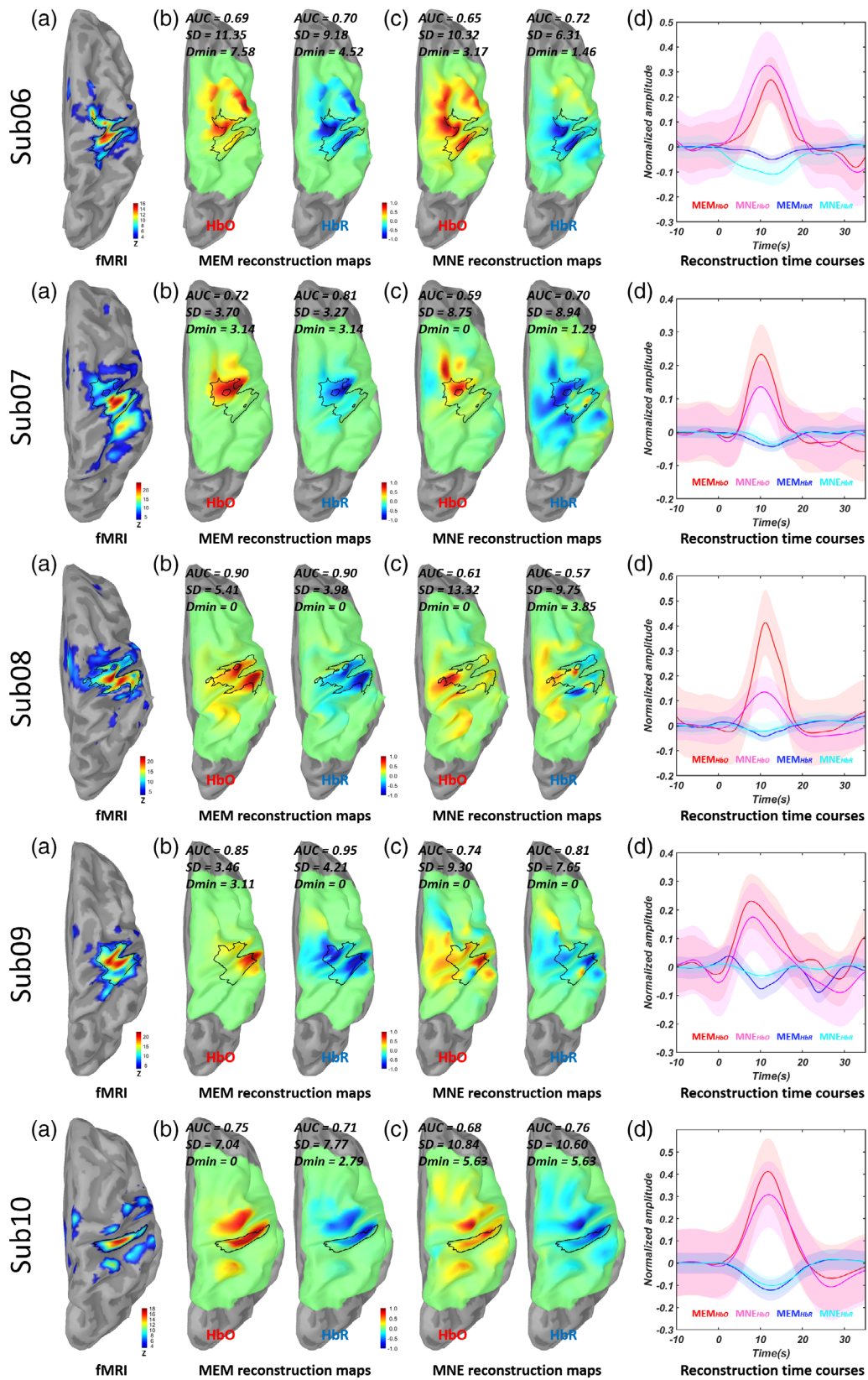


FIGURE 5 Legend on next page.

TABLE 1 Statistical comparison of reconstructed HbO/HbR peak times

	MEM			MNE		
	Peak _{HbO} (s)	Peak _{HbR} (s)	Peak _{HbR-HbO} (s)	Peak _{HbO} (s)	Peak _{HbR} (s)	Peak _{HbR-HbO} (s)
Mean ± se	11.7 ± 0.7	12.7 ± 0.5	0.9 ± 0.3	11.6 ± 0.5	13.5 ± 1.0	1.9 ± 0.7
[Q ₁ , Q ₃]	[11.0, 12.5]	[11.7, 13.3]	[0.1, 2.0]	[10.8, 12.6]	[12.1, 13.8]	[0.8, 2.4]
Mean > 0	—	—	<i>p</i> = .03	—	—	<i>p</i> = .03
MEM vs. MNE	<i>p</i> = .71	<i>p</i> = .17	<i>p</i> = .27	—	—	—

Note: The first row showed the mean ± standard error of each corresponding peak time in seconds. Peak_{HbR-HbO} indicated the time delay between peak time of HbR (Peak_{HbR}) and HbO (Peak_{HbO}). The second row listed the first quartile (Q₁) to third quartile (Q₃) of each peak time and delay. The third row indicated the *p*-value of one sample *t* test against 0 s of peak time delay estimated by MEM and MNE. The last row demonstrated the *p*-value of paired two sample *t* test, when comparing MEM and MNE results, for each peak time and delay. “—” represented the cases in which no statistical test was conducted since we consider it was not necessary to test whether Peak_{HbO} or Peak_{HbR} itself were significantly larger than 0. Red background indicated significant differences with *p* < .05.

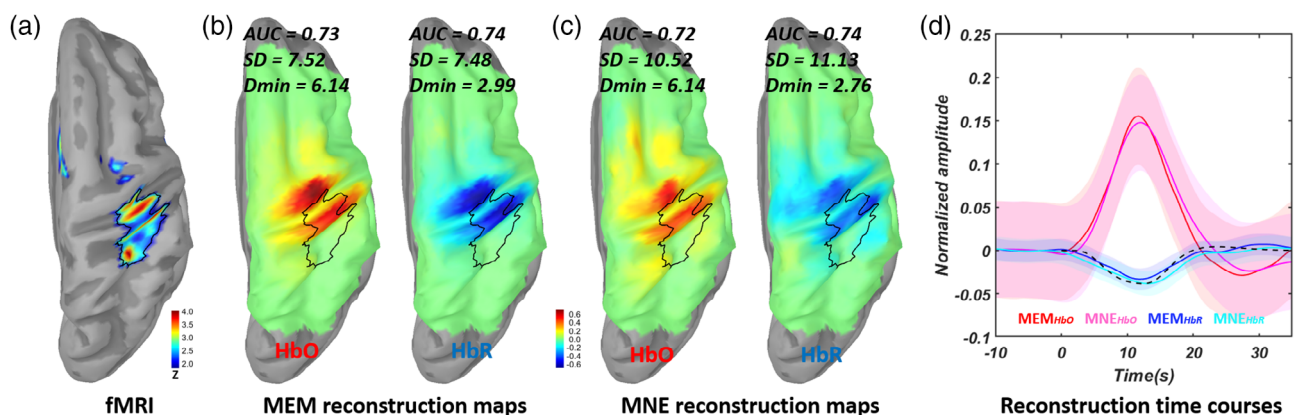


FIGURE 6 Comparisons of MEM and MNE group averaged reconstructions with fMRI Z-maps at the group-level, 10 subjects included. (a) Group-level fMRI Z-maps estimated by FSL group-level activation map projected along the mid-surface of ICBM152 mid-surface. The black profile represents fMRI the most significant cluster projected on the cortical surface. Maps were thresholded to exclude the nonsignificant regions along the cortical surface. (b) Group averaged MEM reconstruction peak maps, individual peak maps were extracted at the peak times of HbO/HbR in native space and then projected onto the ICBM152 mid-surface, (c) group averaged MNE reconstruction peak maps, individual peak maps were extracted at the peak times of HbO/HbR in native space and then projected onto the ICBM152 mid-surface. (d) Group-level reconstructed time courses within the black profile, solid lines represent the averaged time courses, and the shaded areas represent standard deviation within the black profile. The black dash line represents the expected fMRI time course resulting from the convolution of the standard canonical HRF and a 10s boxcar representing the task stimuli. The amplitude was reversed for better comparison with HbR time courses

hemodynamic fluctuations evoked by a motor task. This workflow attempted to optimize the reconstruction accuracy considering (a) fNIRS montage planning using personalized optimal montage (Machado et al., 2014; Machado et al., 2018), which aimed to maximize the fNIRS probing ability of the hemodynamic responses within a targeted ROI along the cortical surface; (b) data acquisition involving

the neuronavigation-guided optode installation using collodion, which ensured accurate positioning, excellent contact to the scalp and minimized motion artifacts (Machado et al., 2018; Pellegrino et al., 2016; Yücel et al., 2014), digitalization of the positions of the sensors along with more than 150 head points, accurate coregistration with the anatomical head model, and therefore minimizing potential errors when

FIGURE 5 Visual and quantitative evaluation for NIROT images at the individual level (subjects 6 to 10). (a) individual level fMRI Z-map thresholded to only contain the significant activation of each subject during finger tapping task, black profile represents the fMRI reference ROI#1 (“true positives”), nonthresholded gray area (nonsignificant) within the reconstruction FOV represents the fMRI reference ROI#2 (“true negatives”). (b) MEM reconstructed NIROT maps at HbO/HbR peak amplitude, respectively, (c) MNE reconstructed NIROT maps at HbO/HbR peak amplitude, respectively, (d) reconstructed time courses within the black profile, solid lines represent the averaged time courses, and the shaded areas represent standard deviation within the ROI#1. Quantitative evaluation metric results are showed on top of each map, respectively. AUCs were calculated by setting black profiles (fMRI reference ROI#1) as the “activated” region and gray area (fMRI reference ROI#2) as the “non-activated” region. SD and Dmin only considered fMRI ROI#1 as the “activated” region. Note that each subject’s map was normalized by its own peak, and time courses were normalized by the maximum HbO amplitude across two methods before averaging within the black profile

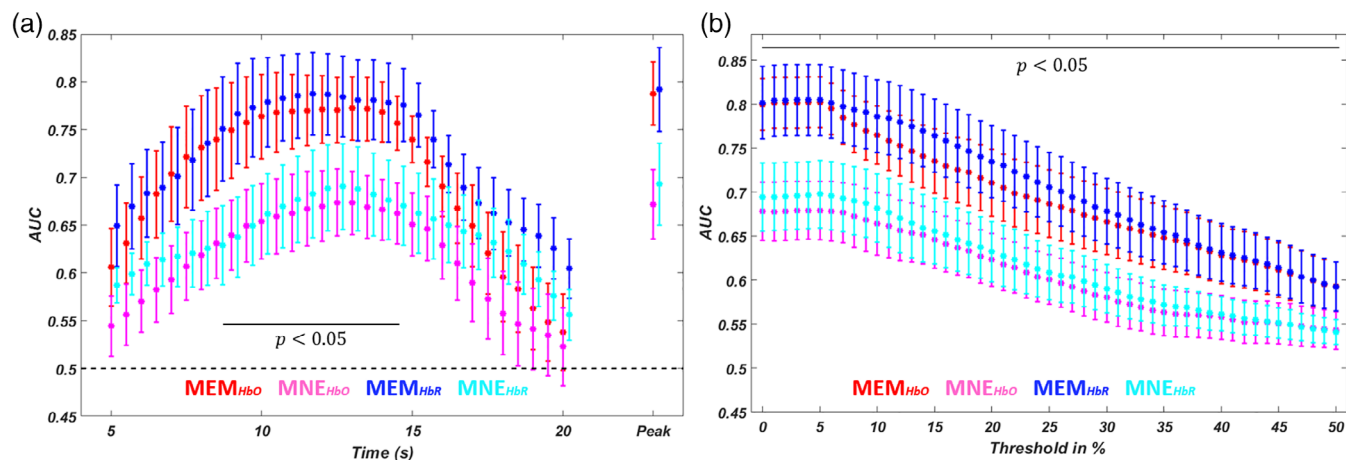


FIGURE 7 AUC comparisons of MEM and MNE reconstructions with respect to time and amplitude threshold value. (a) AUC comparison between MEM and MNE, for both HbO and HbR reconstructions. AUC values (mean \pm standard error) were estimated within the time range 5s – 20s with increments of 0.5s. The last column denoted as “Peak” reports the distribution of AUC values considered at the peak of the hemodynamic response, within the ROI#1, (b) AUC comparison between MEM and MNE, for both HbO and HbR reconstructions, AUC values (mean \pm standard error) were estimated within a percentage threshold of the peak amplitude of each NIROT maps ranging from 0% to 50% with an increment of 1%. Error bars represented the standard error within 10 subjects. Horizontal black bars indicated a significant difference of AUC between MEM and MNE, paired two sample *t* test at each time sample or percentage threshold, $p < .05$ for both HbO and HbR comparisons

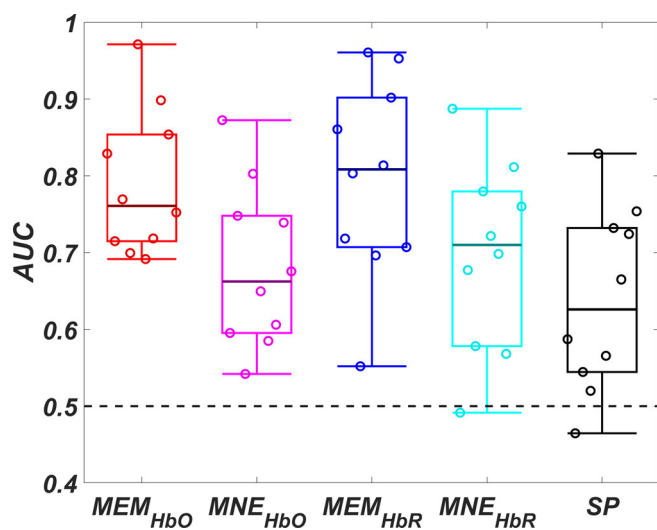


FIGURE 8 Distribution of AUC values estimated for MEM and MNE local reconstructions with the personalized optimal montage, in comparison to the AUC values obtained when considering only the montage light sensitivity profile (SP) for each subject. Exactly the same calculations of AUC for reconstruction methods were applied considering the light sensitivity profile (forward problem) as the reconstructed map. Paired two sample *t* test showed significant higher AUC of MEM reconstructed HbO (mean = 0.76) and HbR (mean = 0.81) than “SP” (mean = 0.64), $p < .01$, Bonferroni corrected for multiple comparison). AUC of “SP” was lower than MNE reconstructed HbO (mean = 0.66) and HbR (mean = 0.71) but not significant

calculating the forward model of NIROT; (c) reconstruction using MEM, which is sensitive to the spatial extent of the generators in the context of EEG/MEG source location (Chowdhury et al., 2013;

Chowdhury et al., 2016; Grova et al., 2016; Hedrich et al., 2017; Heers et al., 2016; Pellegrino et al., 2020) and fNIRS reconstructions with realistic simulations (Cai et al., 2021). We evaluated the spatial accuracy of the NIROT workflow by comparing the HbO/HbR maps to fMRI Z-maps. Our results showed that MEM provided overall better spatial accuracy than MNE, while both NIROT methods exhibited similar temporal features when estimating the fNIRS hemodynamic responses.

4.1 | Evaluation of the performance of personalized NIROT using MEM

Individual and group analysis revealed that MEM recovers more accurately the hemodynamic responses for both HbO and HbR, when compared to MNE (significantly larger AUC). MEM was sensitive to the spatial extent of the generator (significantly lower SD), in agreement with our previous EEG/MEG results (Chowdhury et al., 2013; Chowdhury et al., 2016; Grova et al., 2016; Hedrich et al., 2017; Heers et al., 2016; Pellegrino et al., 2020). Additionally, when assessing the reliability of our measures using discriminability (Wang et al., 2020; Yu, Linn, et al., 2018) through bootstrap resampled sub-averaged task responses, we found that MEM provided reliably better reconstruction performance than MNE. When evaluating the discriminability of MEM and MNE individually, MNE exhibited very large reliability values (e.g., 0.85 for AUC, 0.88 for SD), which were even larger than the known largest reliability values reported in neuroimaging fields—0.8 for morphological measures of the human brain (Zuo et al., 2019). This finding might be explained since MNE is known to spatially smear the reconstruction map, therefore reducing sensitivity and specificity (Ding, 2009), hence resulting in reliable but less

accurate results when compared to MEM. On the other hand, the reliability of MEM alone (0.71 for AUC, 0.66 for SD and 0.64 for Dmin) was overall good.

MEM reconstructed temporal fluctuations of HbO and HbR were similar to MNE (last column of Figures 4 and 5). The statistical analysis of the reconstructed HbO/HbR peak times in Table 1 further proved this point as no significant differences were found between MEM and MNE regarding the HbO/HbR peak times and the delay between them. Moreover, both methods provided significant delay around 1–2 s between HbR and HbO peak times consistent with previous literature (Jaszewski et al., 2003; Steinbrink et al., 2006). In the end, our group level reconstructed HbR time courses (see Figure 6d) estimated by MEM and MNE were almost identical to the expected fMRI time course estimated through the convolution of a 10 s boxcar with the standard canonical HRF. To the best of our knowledge, this is the first study reproducing this aspect by investigating reconstructed time courses rather than measures in the channel space. It is important to note that the time course reconstructed by MNE directly depends on the averaged signal at the channel level as it is a linear projection. Conversely, MEM is a nonlinear technique applied at every time sample, and not originally optimized for the estimation of resulting time courses. The temporal similarity to MNE further validated our previous improvement of MEM (Cai et al., 2021) on temporal accuracy of fNIRS reconstruction.

4.2 | Comparison of MEM and MNE reconstruction performance along time

Most of the spatial accuracy comparisons in this study were performed with the NIROT maps extracted at the HbO/HbR peak. Would MEM outperform MNE at other time points? We also assessed detection accuracy using AUC along the time course of the hemodynamic response, and our results demonstrated that MEM outperformed MNE at any time instance along the elicited hemodynamic response from 5s to 20s after the task onset (Figure 7).

4.3 | Comparison of MEM and MNE reconstruction performance with spatial thresholding

MNE tends to spatially spread the reconstructions out of the “true positive” region due to the use of L2-norm for the regularization when solving the inverse problem (Ding, 2009). Could *post-hoc* thresholding on the reconstruction map achieve a better spatial extension estimation in MNE? As illustrated in Figure 7b, when increasing the initial thresholding in ROC analysis, MNE still provided smaller AUC values than MEM along all percentage thresholding values from 0% to 50%. Therefore, thresholding the MNE map does not solve the typical rate of false positives introduced by the method. While both methods were localizing accurately the maximum of the underlying activity (see Dmin results), MNE did not recover the underlying spatial extent of the active region, whereas MEM did. This result further supported the

study of Ding (2009), which demonstrated that MNE reconstruction cannot retrieve the spatial extent of the underlying generator.

4.4 | Assessment whether the local forward model of personalized optimal montage was biasing NIROT reconstructions

We also carefully assessed if the use of a local forward model (light sensitivity profile map) obtained from an optimal personalized montage targeting a specific ROI, could bias NIROT reconstruction toward this targeted region. We found that the optimal montage maximized the sensitivity with respect to a targeted ROI, but this does not mean that it was “blind” to the other surrounding areas. As shown in Figure 1c, the optimal montage used in this study provided sensitivity covering not only the ROI (“hand knob”) but also surrounding areas such as the pre-central and post-central cortex. To assess this critical issue, we estimated what would be the detection accuracy of the personalized sensitivity profile of each individual optimal montage using AUC (see Figure 1c), assuming that the sensitivity profile itself could be considered as a reconstruction result entirely biased by the optimal montage sensitivity. Our results reported in Figure 8 demonstrated that MEM reconstruction of data from personalized optimal montage provided significantly larger AUC values than the ones estimated when considering only the sensitivity profile as the solution. These results are indeed suggesting that the reconstruction performance is dominated by the process of solving the inverse problem rather than the sensitivity. If optimizing the sensitivity could simply lead the MEM reconstructed activation easily showing up in the targeted ROI, such a clear difference of AUC between sensitivity profile and real reconstructions conducted by MEM should not have been observed. Besides, the apparent difference we found between MEM and MNE reconstruction should not have been observed either. On the other hand, MNE was showing slightly larger AUC when compared to the sensitivity profile (nonsignificant), further demonstrating the inability of the MNE operator to recover accurately the spatial extent. This indicated MNE reconstruction is actually not far from a simple projection of the channel space data onto the cortical surface (Ding, 2009). In the end, these reconstruction results also support our previous conclusion (Machado et al., 2018) that local reconstruction, comparing to NIROT using the high density montage (White & Culver, 2010; Zeff et al., 2007), is possible given appropriate spatial sampling of the montage.

4.5 | Using fMRI as the reference for NIROT evaluation

Using fMRI as the reference to evaluate the accuracy of NIROT reconstructions has been conducted in previous studies (Eggebrecht et al., 2012; Eggebrecht et al., 2014; Huppert, Barker, Schmidt, Walls, & Ghuman, 2017; Tremblay et al., 2018; White & Culver, 2010; Yamashita et al., 2016; Zhan, Eggebrecht, Culver, & Dehghani, 2012).

Mainly three types of reconstruction errors have been considered in these evaluation studies, as follows:

- *The first evaluation criterion* was the overall localization accuracy, as assessed in our study using AUC, also considered in the simulation study (Tremblay et al., 2018) and in a motor task study (Yamashita et al., 2016), which considered both the reconstructed map sensitivity and specificity when compared to the fMRI activation map. Tremblay et al., 2018 used the theoretical ground truth defined in the simulation as the “true positive” region. On the other hand, for real data reconstructions, the calculation of AUC requires the definition of the “true positive” and the “true negative” regions referring to the fMRI activation map. Yamashita et al., 2016 defined the “true positive” region on the fMRI t-map, thresholded at $p < .05$ (Family-wise error rate corrected), and the “true negative” regions with the other voxels exhibiting p values below this threshold. However, in our study, we decided to propose a different approach to determine more robust ROIs from fMRI results as our proposed NIROT validation references.

Our “true positive” region, denoted as fMRI reference ROI#1, consisted in two features, which are respectively size and shape. We proposed a robust approach to define the size of it from the group-level fMRI Z-map for the following reasons: (a) each subject only went through one fMRI task session, and this is not enough to estimate a robust size of the main activation cluster, as suggested by fMRI test-retest reliability studies during motor tasks (Quiton et al., 2014; Zandbelt et al., 2008). (b) Group-level analysis involved more sessions, although collected from different subjects, therefore likely to result in a more robust estimation of the spatial extent of the activated map. This important finding was demonstrated by Zandbelt et al. (2008) who showed a highly stable group-wise spatial activation pattern and BOLD signal changes but substantial variations at the individual level. Then, the shape of the “true positive” region was automatically defined when thresholding the individual fMRI Z-map until the resulted region was exhibiting a similar size to the one defined from the group-level analysis. In the end, the “true positive” region for each individual map was following its own spatial pattern, along the individual level fMRI Z-map, but constrained by a robust cluster size parameter estimated from group-level fMRI analysis. Our resulting threshold informed by the group-level size parameter was usually higher (more conservative) than the conventional threshold considered at the single subject level ($p < .01$, corrected using Gaussian random field theory).

When defining the “true negative” region, denoted as fMRI reference ROI#2, we did not simply use the area that is outside the “true positive” region defined above. Instead, we rather referred to the significance test results of conventional FSL individual voxel-wise fMRI analysis ($p < .01$, corrected using Gaussian random field theory). The nonsignificant fMRI regions projected on the cortical surface were then considered as our “true negative” reference. Our results (Figures 2, 4 and 5) showed that this region was always more extended than the areas outside the “true positive” region. We believe that this proposed approach, discarding the vertices between ROI#1 and ROI#2 from ROC analysis, was fair regarding the evaluation of MNE, since it is known that MNE tends to spatially spread the reconstruction along the cortical region (Ding, 2009). On the other

hand, those “in between” vertices were actually taken into account in our other validation metrics (SD and Dmin).

- *The second evaluation criterion* was the localization error (LE) which can be estimated as suggested in Yamashita et al., 2016 by the distance between the peak of fNIRS reconstruction map and the peak of fMRI Z-map. On the other hand, other studies (Eggebrecht et al., 2012; Eggebrecht et al., 2014; Huppert et al., 2017; White & Culver, 2010) proposed a center of mass error calculated by the distance between the center of mass of two maps (fMRI and NIROT), while others (Tremblay et al., 2018; Zhan et al., 2012) considered the Euclidean distance between reconstruction peak to the peak of the stimulated ground truth. In our study, we defined Dmin as the minimum geodesic distance, following the circumvolutions of the cortical surface, from the vertex that exhibited maximum of reconstructed activity to fMRI reference ROI#1. Overall, we believe that there was little difference between these different localization error metrics, given the fact that reconstructing HbO/HbR peak inside the “true positive” region was not so challenging, resulting in our case of Dmin values mostly close to 0 mm.

- *A third validation criterion* was the false positive or spatial extent reported by the reconstructions. The square root of the area exhibiting amplitudes higher than half maximum, that is, – an estimate of the full width at half maximum (FWHM), was proposed by White and Culver (2010). Similarly, a focality measurement was considered by Zhan et al. (2012), which calculates the ratio between localized full volume half maximum (LVHM) and full volume half maximum (FVHM). FVHM is determined as the volume of all voxels along the whole reconstruction field of view that exhibit amplitude larger than half of the peak amplitude. LVHM is defined similarly but only considering the voxels within the single cluster that contains the peak. Yamashita et al. (2016) estimated a false-positive amount metric defined as the average amount of negative HbR results in the fMRI nonsignificant region, and finally Tremblay et al., 2018 considered a measurement of blurring using the root mean squared of the gradient of the reconstructed map. In our study, we considered the Spatial Dispersion metric (SD), originally proposed by Molins, Stufflebeam, Brown, and Hämäläinen (2008) when studying EEG/MEG source imaging results. SD consists in weighting the distance of the spread of the reconstructed activity around the “true positive” region, by the reconstructed energy in such a region. SD distance in mm is therefore sensitive to how much the reconstructed map is spreading around the “true positive” region but also to eventual mislocalization errors. In our study, by considering SD metric, we are not weighting in a similar manner a false positive activation reported close to the “true positive” region, as opposed to the same amplitude false-positive activation reported far away from the reference region. Therefore, SD is handling correctly such an issue by weighted the distance to the “true positive” region by reconstructed amplitude. When considering only the spatial extent around the peak of the activation, SD also serves similarly as a FWHM measure to quantify the PSF of the reconstructions (Hedrich et al., 2017).

In our previous studies in the context of EEG/MEG source imaging, AUC and SD metrics appeared overall as key metrics to assess

how a source localization or NIROT reconstruction technique could accurately recover the spatial extent of the underlying generators (Cai et al., 2021; Chowdhury et al., 2013; Chowdhury et al., 2016; Hedrich et al., 2017; Pellegrino et al., 2020) therefore they were considered in this study, exhibiting very informative trends in NIROT maps.

4.6 | Difference between NIROT and fMRI maps

Our results also suggested that NIROT images exhibit activations maps more superficial than fMRI Z-maps, both at the individual and group-level analysis. This is in agreement with findings from a simultaneous fMRI/fNIRS study investigating electrical median nerve stimulation (Huppert et al., 2017). This observation is due to the limited ability of fNIRS to detect the deeper areas, a standard limitation of the technique itself (Scholkman et al., 2014) which can be partially compensated by introducing depth weighting when solving the inverse problem. We investigated the impact of depth weighting, in both MEM and MNE, in fNIRS reconstruction in our previous study (Cai et al., 2021). The optimal depth weighting parameters identified in our previous study were actually the ones considered for this study. Therefore, we believe that our reconstructed results should not over-compensating the effect of the depth of the cortical region.

Although fMRI results were considered as our “ground truth,” fMRI and fNIRS signals still have different physical and physiological origins. fMRI measures the Blood-Oxygen-Level-Dependent (BOLD) signal which is a combination of blood flow changes, blood volume changes and deoxygenated hemoglobin concentration changes (Buxton, 2012; Ogawa, Lee, Kay, & Tank, 1990). On the other hand, using infra-red light absorption in two wavelengths, fNIRS measures the relative oxy-/deoxygenated hemoglobin concentration changes evoked by a task or during the resting state (Scholkman et al., 2014). Previous studies have provided conflicting results on whether BOLD signals exhibit the highest correlation with HbO or HbR (Cui, Bray, Bryant, Glover, & Reiss, 2011; Eggebrecht et al., 2012; Eggebrecht et al., 2014; Huppert, Hoge, Diamond, Franceschini, & Boas, 2006; Steinbrink et al., 2006; Strangman, Culver, Thompson, & Boas, 2002; Wijekumar, Huppert, Magnotta, Buss, & Spencer, 2017). HbR is physically closer to fMRI BOLD signal but the correlation is biased by variability resulted from systematic errors (Strangman et al., 2002). We did not fully investigate this aspect, but HbR reconstruction maps showed overall a better consensus with fMRI Z-map (see Figures 4 and 5). For similar reasons, Yamashita et al., 2016 decided to compare only HbR maps in their comparison with fMRI results.

4.7 | Comparison with other NIROT workflows

Most references reported on NIROT results were performed in the context of high-density fNIRS montage (Eggebrecht et al., 2014; Zeff et al., 2007). Increasing the density of the fNIRS montage is expected to improve the power of the reconstruction accuracy and the resolution (White & Culver, 2010). Along with our previous work (Cai

et al., 2021; Machado et al., 2018) and the present study, we showed that our approach allows accurate local reconstruction of NIROT images even when considering fewer well-positioned channels. Moreover, we believe that optimal montage can also be beneficial for studies which require more on portability.

Other NIROT strategies have also been reported in the literature. Linear approaches using L1-norm based regularization (Kavuri, Lin, Tian, & Liu, 2012; Okawa, Hoshi, & Yamada, 2011; Prakash, Shaw, Manjappa, Kanhirojan, & Yalavarthy, 2014; Süzen, Giannoula, & Durduran, 2010) have been used to obtain more focal localization with high specificity but with much lower sensitivity. Bayesian model averaging (Tremblay et al., 2018) proposed by Trujillo-Barreto, Aubert-Vázquez, and Valdés-Sosa (2004) and a hierarchical Bayesian model applied by Shimokawa et al. (2012), Shimokawa et al. (2013), Yamashita et al. (2016) reported more accurate sparse reconstructions with both higher sensitivity and specificity than L1-and L2-norm based regularizations. It would be of great interest to compare our MEM framework with these other NIROT Bayesian methods, as we previously completed in the context of MEG source imaging (Chowdhury et al., 2013), either on realistic simulations or on well controlled experimental data, however such analysis was out of the scope of this study.

Regarding the statistical analysis of NIROT maps, a hierarchical random-effects cortical surface reconstruction model proposed by Abdelnour, Genovese, and Huppert (2010) was applied by Huppert et al. (2017) to estimate a group level hemodynamic responses evoked by parametric median nerve stimulations. Individual channel space optical density changes and forward model (sensitivity) were concatenated into a single inverse model so that the group-level inference can be made directly by solving this model. In our study, we mainly focused on the paired comparison of the reconstruction performance between MEM and MNE at the individual-level, not on statistical inferences at the group-level. However, in addition to standard evaluation of the grand average group-level responses (average of individual NIROT maps) presented in Figure 6, we also investigated group-level evaluation when considering group-level Z maps of MEM and MNE reconstructions (see Figure S2). In this case, the Z score was estimated for each vertex along the cortical surface among 10 subjects' reconstruction maps, using the mean of the normalized amplitude divided by the standard deviation of 10 subjects, therefore resulting in a random effect group-level analysis. Our results further demonstrated a good estimation of the main activation cluster using MEM, when compared to MNE.

4.8 | Implementation of the workflow in an open-source toolbox

Our personalized NIROT workflow is publicly available. All methods described in this study have been implemented in an fNIRS processing plugin—NIRSTORM (<https://github.com/Nirstorm/nirstorm>) in Brainstorm software. Brainstorm (Tadel et al., 2011) (<https://neuroimage.usc.edu/brainstorm/>) is an open-source software dedicated to

analyzing multimodality brain recordings such as MEG, EEG, fNIRS, electrocorticography, depth electrodes and multiunit electrophysiology. Researchers could apply our proposed NIROT workflow in NIRSTORM starting with designing a personalized optimal montage using individual MRI or a template MRI (Colin 27). If gluing optodes on any possible location is not feasible, our optimal montage can also be estimated on a discrete set of free optodes positions available on a specific cap. fNIRS data preprocessing using the conventional pipeline, and finally reconstruction NIROT spatiotemporal images using depth weighted versions of MEM and MNE are all available in NIRSTORM.

4.9 | Limitations and future directions

The subject cohort involved in this study was still relatively small—10 subjects. This could reduce the power of our statistical analysis. However, the design of the study was fully within-subject, with paired comparisons which typically grants more statistical power. The input to each reconstruction method contained the exact same variances of hemodynamic responses between- and within-subjects, meaning the paired comparisons or contrasts involved in this study are not biased by such variances. When studying the reliability of fNIRS 3D reconstruction performances, within-subject variability was conducted using a resampling sub-averaged approach. Although it was beyond the scope of this study, it would be of great interest to perform such reliability evaluation with a well-designed test–retest reliability study. It is worth noting that a recent fNIRS reproducibility study has demonstrated the importance of involving montage spatial information, which specifically increased the within-subject reproducibility (Novi et al., 2020), hence taking advantages of the personalized optimal montage our proposed workflow is expected to provide high reliability in a test–retest reliability evaluation study.

Although the same task was performed by the same cohort, fMRI and fNIRS acquisitions were not conducted simultaneously. Within-subject task performance variability might also cause some potential differences between the NIROT and fMRI images, including influence of other processes like attention or arousal (Novi et al., 2020). Concurrent fMRI/fNIRS acquisitions (Huppert et al., 2017; Wijekumar et al., 2017) may be used to address this problem but this was beyond the scope of this study. Overall, we believe there should be great interest to apply the proposed NIROT workflow into different experiments designs, involving a variety of tasks.

5 | CONCLUSION

In the present study, we demonstrated and evaluated our proposed workflow for personalized NIROT using MEM. Finger tapping data acquired from 10 subjects were used to reconstruct NIROT images and validating the results by comparing them to the fMRI Z-maps obtained from the same task and cohort. Conventionally used reconstruction method—MNE was compared with MEM. Our results showed that MEM provides better spatial accuracy and similar

temporal features as compared with MNE, when reconstructing both HbO and HbR responses evoked by a finger tapping task. Our workflow is publicly available, and all the processing functionalities have been implemented and validated in the fNIRS processing plugin—NIRSTORM (<https://github.com/Nirstorm/nirstorm>) in Brainstorm software.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant Program (CG and JML), an operating grant from the Canadian Institutes for Health Research (CIHR MOP 133619 [CG]) and a FRQNT research team grant. fNIRS equipment was acquired using grants from NSERC Research Tools and Instrumentation Program and the Canadian Foundation for Innovation (CG). ZC is funded by the Fonds de recherche du Québec – Santé (FRQS) Doctoral Training Scholarship and the PERFORM Graduate Scholarship in Preventive Health Research. GP is funded by Strauss Canada Foundation and a McGill/MNI—Fred Andermann EEG and Epilepsy Fellowship. MU is funded by the Horizon Postdoctoral Fellowships of Concordia University. UA was funded by the FRQS postdoctoral fellowship and Savoy Foundation postdoctoral fellowship.

CONFLICT OF INTERESTS

The authors declare no potential conflict of interest.

DATA AVAILABILITY STATEMENT

The original raw data supporting the findings of this study are available upon reasonable request to the corresponding authors.

ORCID

Zhengchen Cai  <https://orcid.org/0000-0002-4233-5568>

Makoto Uji  <https://orcid.org/0000-0002-9445-6353>

Giovanni Pellegrino  <https://orcid.org/0000-0002-1195-1421>

REFERENCES

- Aarabi, A., Osharina, V., & Wallois, F. (2017). Effect of confounding variables on hemodynamic response function estimation using averaging and deconvolution analysis: An event-related NIRS study. *NeuroImage*, 155, 25–49.
- Abdelnour, F., Genovese, C., & Huppert, T. (2010). Hierarchical Bayesian regularization of reconstructions for diffuse optical tomography using multiple priors. *Biomedical Optics Express*, 1, 1084–1103.
- Amblard, C., Lapalme, E., & Lina, J. M. (2004). Biomagnetic source detection by maximum entropy and graphical models. *IEEE Transactions on Biomedical Engineering*, 51, 427–442.
- Arridge, S. R. (1999). Optical tomography in medical imaging. *Inverse Problems*, 15, R41–R93.
- Avigdor, T., Abdallah, C., von Ellenrieder, N., Hedrich, T., Rubino, A., Lo, R. G., ... Frauscher, B. (2021). Fast oscillations >40 Hz localize the epileptogenic zone: An electrical source imaging study using high-density electroencephalography. *Clinical Neurophysiology*, 132(2), 568–580. <https://doi.org/10.1016/j.clinph.2020.11.031>.
- Aydin, Ü., Pellegrino, G., Ali, O. B. K., Abdallah, C., Abdallah, C., Dubeau, F., ... Grova, C. (2020). Magnetoencephalography resting state connectivity patterns as indicatives of surgical outcome in epilepsy patients. *Journal of Neural Engineering*, 17, 35007.

- Boas, D. A., Brooks, D. H., Miller, E. L., Dimarzio, C. A., Kilmer, M., Gaudette, R. J., & Zhang, Q. (2001). Imaging the body with diffuse optical tomography. *IEEE Signal Processing Magazine*, 18, 57–75.
- Boas, D. A., & Dale, A. M. (2005). Simulation study of magnetic resonance imaging-guided cortically constrained diffuse optical tomography of human brain function. *Applied Optics*, 44, 1957–1968.
- Boas, D. A., Gaudette, T., Strangman, G., Cheng, X., Marota, J. J. A., & Mandeville, J. B. (2001). The accuracy of near infrared spectroscopy and imaging during focal changes in cerebral hemodynamics. *NeuroImage*, 13, 76–90.
- Bridgeford, E., Shen, C., Wang, S., & Vogelstein, J. (2018). Multiscale graph correlation. <https://zenodo.org/record/1246967>
- Buxton, R. B. (2012). Dynamic models of BOLD contrast. *NeuroImage*, 62, 953–961.
- Cai, Z., Machado, A., Chowdhury, R. A., Spilkin, A., Vincent, T., Aydin, U., ... Grova, C. (2021). Diffuse optical reconstructions of NIRS data using maximum entropy on the mean. [bioRxiv:2021.02.22.432263](https://arxiv.org/abs/2021.02.22.432263).
- Chowdhury, R. A., Lina, J. M., Kobayashi, E., & Grova, C. (2013). MEG source localization of spatially extended generators of epileptic activity: Comparing entropic and hierarchical Bayesian approaches. *PLoS One*, 8(2), e55969.
- Chowdhury, R. A., Merlet, I., Birot, G., Kobayashi, E., Nica, A., Biraben, A., ... Grova, C. (2016). Complex patterns of spatially extended generators of epileptic activity: Comparison of source localization methods cMEM and 4-ExSo-MUSIC on high resolution EEG and MEG data. *NeuroImage*, 143, 175–195.
- Chowdhury, R. A., Pellegrino, G., Aydin, Ü., Lina, J. M., Dubeau, F., Kobayashi, E., & Grova, C. (2018). Reproducibility of EEG-MEG fusion source analysis of interictal spikes: Relevance in presurgical evaluation of epilepsy. *Human Brain Mapping*, 39, 880–901.
- Cui, X., Bray, S., Bryant, D. M., Glover, G. H., & Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *NeuroImage*, 54, 2808–2821.
- Dehghani, H., White, B. R., Zeff, B. W., Tizzard, A., & Culver, J. P. (2009). Depth sensitivity and image reconstruction analysis of dense imaging arrays for mapping brain function with diffuse optical tomography. *Applied Optics*, 48, D137–D143.
- Delpy, D. T., Cope, M., Van Der Zee, P., Arridge, S., Wray, S., & Wyatt, J. (1988). Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in Medicine and Biology*, 33, 1433–1442.
- Ding, L. (2009). Reconstructing cortical current density by exploring sparseness in the transform domain. *Physics in Medicine and Biology*, 54, 2683–2697.
- Eggebrecht, A. T., Ferradal, S. L., Robichaux-Viehoever, A., Hassanpour, M. S., Dehghani, H., Snyder, A. Z., ... Culver, J. P. (2014). Mapping distributed brain function and networks with diffuse optical tomography. *Nature Photonics*, 8, 448–454.
- Eggebrecht, A. T., White, B. R., Ferradal, S. L., Chen, C., Zhan, Y., Snyder, A. Z., ... Culver, J. P. (2012). A quantitative spatial comparison of high-density diffuse optical tomography and fMRI cortical mapping. *NeuroImage*, 61, 1120–1128.
- Fang, Q., & Boas, D. A. (2009). Monte Carlo simulation of photon migration in 3D turbid media accelerated by graphics processing units. *Optics Express*, 17, 20178–20190.
- Ferradal, S. L., Liao, S. M., Eggebrecht, A. T., Shimony, J. S., Inder, T. E., Culver, J. P., & Smyser, C. D. (2016). Functional imaging of the developing brain at the bedside using diffuse optical tomography. *Cerebral Cortex*, 26, 1558–1568.
- Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 11050–11055.
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., ... Dale, A. M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33, 341–355.
- Fishell, A. K., Burns-Yocum, T. M., Bergonzi, K. M., Eggebrecht, A. T., & Culver, J. P. (2019). Mapping brain function during naturalistic viewing using high-density diffuse optical tomography. *Scientific Reports*, 9, 11115.
- Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., ... Mattout, J. (2008). Multiple sparse priors for the M/EEG inverse problem. *NeuroImage*, 39, 1104–1120.
- Grova, C., Aiguabella, M., Zelmann, R., Lina, J. M., Hall, J. A., & Kobayashi, E. (2016). Intracranial EEG potentials estimated from MEG sources: A new approach to correlate MEG and iEEG data in epilepsy. *Human Brain Mapping*, 37, 1661–1683.
- Grova, C., Daunizeau, J., Lina, J. M., Bénar, C. G., Benali, H., & Gotman, J. (2006). Evaluation of EEG localization methods using realistic simulations of interictal spikes. *NeuroImage*, 29, 734–753.
- Grova, C., Makni, S., Flandin, G., Ciuciu, P., Gotman, J., & Poline, J. B. (2006). Anatomically informed interpolation of fMRI data on the cortical surface. *NeuroImage*, 31, 1475–1486.
- Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & Biological Engineering & Computing*, 32, 35–42.
- Hansen, P. C. (2000). The L-curve and its use in the numerical treatment of inverse problems. *Computational Inverse Problems in Electrocardiology. Advances in Computational Bioengineering*, 4, 119–142.
- Hassanpour, M. S., Eggebrecht, A. T., Peelle, J. E., & Culver, J. P. (2017). Mapping effective connectivity within cortical networks with diffuse optical tomography. *Neurophotonics*, 4, 041402.
- Hedrich, T., Pellegrino, G., Kobayashi, E., Lina, J. M., & Grova, C. (2017). Comparison of the spatial resolution of source imaging techniques in high-density EEG and MEG. *NeuroImage*, 157, 531–544.
- Heers, M., Chowdhury, R. A., Hedrich, T., Dubeau, F., Hall, J. A., Lina, J. M., ... Kobayashi, E. (2016). Localization accuracy of distributed inverse solutions for electric and magnetic source imaging of Interictal epileptic discharges in patients with focal epilepsy. *Brain Topography*, 29, 162–181.
- Huppert, T., Barker, J., Schmidt, B., Walls, S., & Ghuman, A. (2017). Comparison of group-level, source localized activity for simultaneous functional near-infrared spectroscopy-magnetoencephalography and simultaneous fNIRS-fMRI during parametric median nerve stimulation. *Neurophotonics*, 4, 015001.
- Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage*, 29, 368–382.
- Jasdzewski, G., Strangman, G., Wagner, J., Kwong, K. K., Poldrack, R. A., & Boas, D. A. (2003). Differences in the hemodynamic response to event-related motor and visual paradigms as measured by near-infrared spectroscopy. *NeuroImage*, 20, 479–488.
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, 62, 782–790.
- Jöbsis, F. F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science (80-)*, 198, 1264–1266.
- Kavuri, V. C., Lin, Z.-J., Tian, F., & Liu, H. (2012). Sparsity enhanced spatial resolution and depth localization in diffuse optical tomography. *Bio-medical Optics Express*, 3, 943–957.
- Lapalme, E., Lina, J. M., & Mattout, J. (2006). Data-driven parceling and entropic inference in MEG. *NeuroImage*, 30, 160–171.
- Machado, A., Cai, Z., Pellegrino, G., Marcotte, O., Vincent, T., Lina, J.-M., ... Grova, C. (2018). Optimal positioning of optodes on the scalp for personalized functional near-infrared spectroscopy investigations. *Journal of Neuroscience Methods*, 309, 91–108.
- Machado, A., Marcotte, O., Lina, J. M., Kobayashi, E., & Grova, C. (2014). Optimal optode montage on electroencephalography/functional near-

- infrared spectroscopy caps dedicated to study epileptic discharges. *Journal of Biomedical Optics*, 19, 026010.
- Mattout, J., Pélégriani-Issac, M., Garnero, L., & Benali, H. (2005). Multivariate source prelocalization (MSP): Use of functionally informed basis functions for better conditioning the MEG inverse problem. *NeuroImage*, 26, 356–373.
- Molins, A., Stufflebeam, S. M., Brown, E. N., & Hämäläinen, M. S. (2008). Quantification of the benefit from integrating MEG and EEG data in minimum ℓ_2 -norm estimation. *NeuroImage*, 42, 1069–1077.
- Novi, S. L., Forero, E. J., Rubianes Silva, J. A. I., de Souza, N. G. S. R., Martins, G. G., Quiroga, A., ... Mesquita, R. C. (2020). Integration of spatial information increases reproducibility in functional near-infrared spectroscopy. *Frontiers in Neuroscience*, 14, 1–12.
- Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87, 9868–9872.
- Okawa, S., Hoshi, Y., & Yamada, Y. (2011). Improvement of image quality of time-domain diffuse optical tomography with LP sparsity regularization. *Biomedical Optics Express*, 2, 3334–3348.
- Pellegrino, G., Hedrich, T., Chowdhury, R., Hall, J. A., Lina, J. M., Dubeau, F., ... Grova, C. (2016). Source localization of the seizure onset zone from ictal EEG/MEG data. *Human Brain Mapping*, 37, 2528–2546.
- Pellegrino, G., Hedrich, T., Porras-Bettancourt, M., Lina, J. M., Aydin, Ü., Hall, J., ... Kobayashi, E. (2020). Accuracy and spatial properties of distributed magnetic source imaging techniques in the investigation of focal epilepsy patients. *Human Brain Mapping*, 41, 3019–3033.
- Pellegrino, G., Machado, A., von Ellenrieder, N., Watanabe, S., Hall, J. A., Lina, J. M., ... Grova, C. (2016). Hemodynamic response to interictal epileptiform discharges addressed by personalized EEG-fNIRS recordings. *Frontiers in Neuroscience*, 10, 102.
- Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical parametric mapping: The analysis of functional brain images*. Amsterdam, The Netherlands: Elsevier.
- Piper, S. K., Krueger, A., Koch, S. P., Mehnert, J., Habermehl, C., Steinbrink, J., ... Schmitz, C. H. (2014). A wearable multi-channel fNIRS system for brain imaging in freely moving subjects. *NeuroImage*, 85, 64–71.
- Prakash, J., Shaw, C. B., Manjappa, R., Kanhiroan, R., & Yalavarthy, P. K. (2014). Sparse recovery methods hold promise for diffuse optical tomographic image reconstruction. *IEEE Journal of Selected Topics in Quantum Electronics*, 20, 74–82.
- Qutun, R. L., Keaser, M. L., Zhuo, J., Gullapalli, R. P., & Greenspan, J. D. (2014). Intersession reliability of fMRI activation for heat pain and motor tasks. *NeuroImage: Clinical*, 5, 309–321.
- Raffin, E., Pellegrino, G., Di Lazzaro, V., Thielscher, A., & Siebner, H. R. (2015). Bringing transcranial mapping into shape: Sulcus-aligned mapping captures motor somatotopy in human primary motor hand area. *NeuroImage*, 120, 164–175.
- Schmitz, C. H., Klemer, D. P., Hardin, R., Katz, M. S., Pei, Y., Graber, H. L., ... Barbour, R. L. (2005). Design and implementation of dynamic near-infrared optical tomographic imaging instrumentation for simultaneous dual-breast measurements. *Applied Optics*, 44, 2140–2153.
- Schneider, P., Piper, S., Schmitz, C. H., Schreiter, N. F., Volkwein, N., Lüdemann, L., ... Poellinger, A. (2011). Fast 3D near-infrared breast imaging using indocyanine green for detection and characterization of breast lesions. *RoFo Fortschritte auf dem Gebiet der Röntgenstrahlen und der Bildgeb Verfahren*, 183, 956–963.
- Scholkman, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., & Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *NeuroImage*, 85, 6–27.
- Shimokawa, T., Kosaka, T., Yamashita, O., Hiroe, N., Amita, T., Inoue, Y., & Sato, M. (2012). Hierarchical Bayesian estimation improves depth accuracy and spatial resolution of diffuse optical tomography. *Optics Express*, 20, 20427–20446.
- Shimokawa, T., Kosaka, T., Yamashita, O., Hiroe, N., Amita, T., Inoue, Y., & Sato, M. (2013). Extended hierarchical Bayesian diffuse optical tomography for removing scalp artifact. *Biomedical Optics Express*, 4, 2411–2432.
- Steinbrink, J., Villringer, A., Kempf, F., Haux, D., Boden, S., & Obrig, H. (2006). Illuminating the BOLD signal: Combined fMRI-fNIRS studies. *Magnetic Resonance Imaging*, 24, 495–505.
- Strangman, G., Culver, J. P., Thompson, J. H., & Boas, D. A. (2002). A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. *NeuroImage*, 17, 719–731.
- Strangman, G., Franceschini, M. A., & Boas, D. A. (2003). Factors affecting the accuracy of near-infrared spectroscopy concentration calculations for focal changes in oxygenation parameters. *NeuroImage*, 18, 865–879.
- Süzen, M., Giannoula, A., & Durduran, T. (2010). Compressed sensing in diffuse optical tomography. *Optics Express*, 18, 23676–23690.
- Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, 2011, 879716.
- Tremblay, J., Martínez-Montes, E., Vannasing, P., Nguyen, D. K., Sawan, M., Lepore, F., & Gallagher, A. (2018). Comparison of source localization techniques in diffuse optical tomography for fNIRS application using a realistic head model. *Biomedical Optics Express*, 9, 2994–3016.
- Trujillo-Barreto, N. J., Aubert-Vázquez, E., & Valdés-Sosa, P. A. (2004). Bayesian model averaging in EEG/MEG imaging. *NeuroImage*, 21, 1300–1319.
- Vogelstein, J. T., Bridgeford, E. W., Wang, Q., Priebe, C. E., Maggioni, M., & Shen, C. (2019). Discovering and deciphering relationships across disparate data modalities. *eLife*, 8, 1–32.
- Wang, Z., Bridgeford, E., Wang, S., Vogelstein, J. T., & Caffo, B. (2020). Statistical analysis of data repeatability measures. arXiv:1–26.
- White, B. R., & Culver, J. P. (2010). Quantitative evaluation of high-density diffuse optical tomography: In vivo resolution and mapping performance. *Journal of Biomedical Optics*, 15, 026006.
- Wijeakumar, S., Huppert, T. J., Magnotta, V. A., Buss, A. T., & Spencer, J. P. (2017). Validating an image-based fNIRS approach with fMRI and a working memory task. *NeuroImage*, 147, 204–218.
- Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for fMRI group analysis using Bayesian inference. *NeuroImage*, 21, 1732–1747.
- Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of fMRI data. *NeuroImage*, 14, 1370–1386.
- Worsley, K. J. (2001). In P. Jezzard, P. M. Matthews, & S. M. Smith (Eds.), *Statistical analysis of activation images. Ch 14, in functional MRI: An introduction to methods*. New York: Oxford University Press.
- Yamashita, O., Shimokawa, T., Aisu, R., Amita, T., Inoue, Y., & Sato, M. (2016). Multi-subject and multi-task experimental validation of the hierarchical Bayesian diffuse optical tomography algorithm. *NeuroImage*, 135, 287–299.
- Yu, L., Nina-Paravecino, F., Kaeli, D., & Fang, Q. (2018). Scalable and massively parallel Monte Carlo photon transport simulations for heterogeneous computing platforms. *Journal of Biomedical Optics*, 23, 1.
- Yu, M., Linn, K. A., Cook, P. A., Phillips, M. L., McInnis, M., Fava, M., ... Sheline, Y. I. (2018). Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. *Human Brain Mapping*, 39, 4213–4227.
- Yücel, M. A., Lühmann, A. V., Scholkman, F., Gervain, J., Dan, I., Ayaz, H., ... Wolf, M. (2021). Best practices for fNIRS publications. *Neurophotonics*, 8, 1–34.
- Yücel, M. A., Selb, J., Boas, D. A., Cash, S. S., & Cooper, R. J. (2014). Reducing motion artifacts for long-term clinical NIRS monitoring using collodion-fixed prism-based optical fibers. *NeuroImage*, 85, 192–201.

- Zandbelt, B. B., Gladwin, T. E., Raemaekers, M., van Buuren, M., Neggers, S. F., Kahn, R. S., ... Vink, M. (2008). Within-subject variation in BOLD-fMRI signal changes across repeated measurements: Quantification and implications for sample size. *NeuroImage*, 42, 196–206.
- Zeff, B. W., White, B. R., Dehghani, H., Schlaggar, B. L., & Culver, J. P. (2007). Retinotopic mapping of adult human visual cortex with high-density diffuse optical tomography. *Proceedings of the National Academy of Sciences*, 104, 12169–12174.
- Zhan, Y., Eggebrecht, A. T., Culver, J. P., & Dehghani, H. (2012). Image quality analysis of high-density diffuse optical tomography incorporating a subject-specific head model. *Frontiers in Neuroenergetics*, 4, 6.
- Zuo, X.-N., Xu, T., Milham, M. P. (2019). Harnessing reliability for neuroscience research. *Nature Human Behaviour*, 3(8), 768–771. <https://doi.org/10.1038/s41562-019-0655-x>.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Cai, Z., Uji, M., Aydin, Ü., Pellegrino, G., Spilkin, A., Delaire, É., Abdallah, C., Lina, J.-M., & Grova, C. (2021). Evaluation of a personalized functional near infra-red optical tomography workflow using maximum entropy on the mean. *Human Brain Mapping*, 42(15), 4823–4843. <https://doi.org/10.1002/hbm.25566>

APPENDIX A: fNIRS reconstruction using minimum norm estimate (MNE) and maximum entropy on the mean (MEM)

NIROT reconstruction can be modeled as a linear problem, $Y = AX + e$, where Y ($p \times t$) is the preprocessed ΔOD at a specific wavelength for a specific channel p at a time sample t ; A is the sensitivity matrix ($p \times q$) estimated when solving NIROT forward problem between the vertex q and the channel p . X is a ($q \times t$) matrix representing the reconstructed amplitude at vertex q along the cortical surface at time t , and e is the reconstruction error (same dimension as Y). Solving X by knowing Y and A involves solving an ill-posed inverse problem, which requires regularization. In the MNE approach (Hämäläinen & Ilmoniemi, 1994), Tikhonov regularization was used to minimize the L2-norm, thus the estimated reconstruction spatial-temporal matrix \hat{X}_{MNE} as,

$$\hat{X}_{MNE} = \operatorname{argmin} \left(\|Y - AX\|_{\Sigma_d}^2 + \lambda \|X\|_{\Sigma_s}^2 \right) = (A^T \Sigma_d A + \lambda \Sigma_s)^{-1} A^T \Sigma_d Y$$

where Σ_d and Σ_s are the inverse of noise covariance and source covariance, respectively, λ is the hyperparameter to regularize the inversion. Σ_d was estimated as a full noise covariance matrix from baseline recordings (i.e., -10 s to 0 s). Σ_s is assumed to be an identity matrix in conventional MNE. In our implementation, λ was estimated by the standard L-Curve method, as suggested in Hansen (2000).

In our proposed MEM method, which is a probabilistic framework, the probability distribution of the amplitude of X , described as $dp(x) = p(x)dx$, can be estimated by Bayesian inference, starting from

a predefined prior distribution of X denoted $d\nu(x)$. The peak of the posterior of $dp(x)$ represented by $dp^*(x)$ is estimated by maximizing the Kullback–Leibler divergence or ν -entropy to the prior as following,

$$S_\nu(dp(x)) = - \int \log \left(\frac{dp(x)}{d\nu(x)} \right) dp(x) = - \int f(x) \log(f(x)) d\nu(x)$$

$$dp^*(x) = \operatorname{argmax}_{dp(x) \in C_m} (S_\nu(dp(x)))$$

where $S_\nu(dp(x))$ is the ν -entropy of $dp(x)$ to prior $d\nu(x)$, C_m is the set of probability distributions of x that explains the data Y on average, meaning

$$Y - \left[A \middle| I \right] \begin{bmatrix} E_{dp}[x] \\ e \end{bmatrix} = 0, dp \in C_m$$

where $E_{dp}[x]$ is the statistical expectation of x under the probability distribution dp , I is the identity matrix with the dimension of the number of vertices involved in the reconstruction.

Then, assuming that brain activity could be described by K non-overlapping and independent cortical parcels, we proposed the following reference distribution $d\nu(x)$ model,

$$d\nu(x) = \prod_{k=1}^K [(1 - \alpha_k) \delta(x_k) + \alpha_k N(\mu_k, \Sigma_k)] dx_k, \quad 0 < \alpha_k < 1$$

where the hidden variable S_k defines the activation state (active or not) of each cortical parcel k . α_k is the probability of k^{th} parcel to be active, that is, $\operatorname{Prob}(S_k = 1)$. δ_k is a Dirac function that allows to “switch off” the parcel when considered as inactive (i.e., $S_k = 0$). $N(\mu_k, \Sigma_k)$ is a Gaussian distribution, describing the distribution of absorptions changes within the k^{th} parcel, when the parcel is considered as active, $S_k = 1$. Note that the multiplication in the definition of $d\nu(x)$ is referring to the assumption that all parcels are statistically independent.

A data driven parcellation (DDP) technique (Lapalme, Lina, & Mattout, 2006) was used to parcellate the cortical surface into K nonoverlapping parcel. The probability of each parcel to be active (α_k) was initialized as the median multivariate source prelocalization (MSP) (Mattout, Pélégriani-Issac, Garnero, & Benali, 2005) score from all the sources within the parcels.

To initialize the $N(\mu_k, \Sigma_k)$ in prior $d\nu(x)$, μ_k was set to zero. $\Sigma_k(t)$ at each time point t was defined according to Chowdhury et al. (2013),

$$\Sigma_k(t) = \eta(t) W_k(\sigma)^T W_k(\sigma)$$

$$\eta(t) = 0.05 \frac{1}{P_k} \sum_{i \in P_k} \hat{X}_{MNE}^2(i, t)$$

where $W_k(\sigma)$ is a spatial smoothness matrix, defined by Friston et al. (2008), which controls the local spatial smoothness within the parcel according to the geodesic surface neighborhood order. $\eta(t)$ was defined as 5% of the averaged energy of MNE solution within each parcel.

Note that for NIROT reconstruction, we also applied a depth-weighted version of MEM and MNE, as described and evaluated in Cai et al. (2021). Please refer to this work for further methodological details.

APPENDIX B: The reliability of performance differences between MEM and MNE

FIGURE B1 Histograms of paired evaluation metric differences between MEM and MNE estimated over 100 sessions (10 within subject resampled sessions \times 10 subjects). The blue dashed lines indicated the 95% confidence interval (CI) of the estimated paired differences on (a) AUC, (b) SD (in mm) and (c) Dmin (in mm). Red dashed lines showed 0 difference and black dashed lines represented the mean of the corresponding differences. Based on all sessions including within- and between-subjects variability, MEM provided significantly higher AUCs (overall better sensitivity and specificity), smaller SDs (less amplitude weighted spatial spread) and smaller Dmin (closer to the main fMRI cluster) than MNE

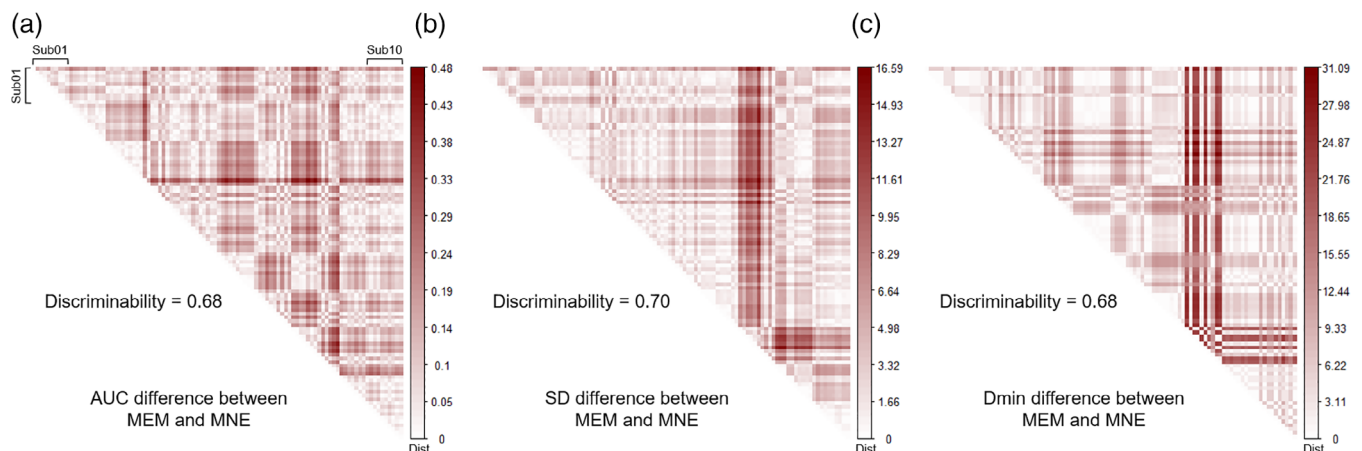
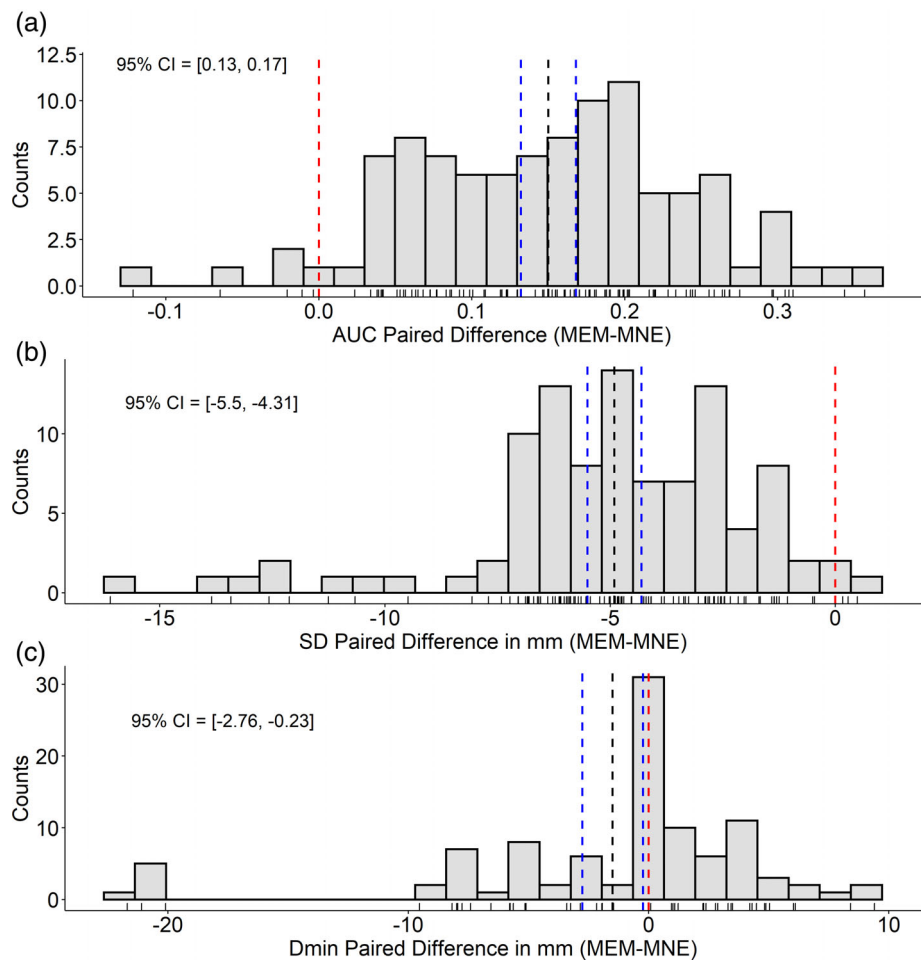


FIGURE B2 Evaluations of reliability of performance differences between MEM and MNE using discriminability. Each diagonal matrix demonstrated the within-subject and between-subject Euclidean distances of the pair-wised performance differences (MEM – MNE) for (a) AUC, (b) SD (in mm) and (c) Dmin (in mm). MEM showed reliably better reconstruction performance than MNE indicated by a discriminability measure of 0.68 for AUC, 0.70 for SD and 0.68 for Dmin, as illustrated in the figure showing that within-subject distances (block diagonal terms) were overall smaller than between-subject distances (off diagonal terms)