



HHS Public Access

Author manuscript

IEEE Trans Neural Syst Rehabil Eng. Author manuscript; available in PMC 2021 September 20.

Published in final edited form as:

IEEE Trans Neural Syst Rehabil Eng. 2021 ; 29: 1679–1689. doi:10.1109/TNSRE.2021.3105644.

Efficient TMS-Based Motor Cortex Mapping Using Gaussian Process Active Learning

Razieh Faghihpirayesh,

Cognitive Systems Laboratory, Electrical and Computer Engineering Department, Northeastern University, Boston, MA 02115 USA; Center for Signal Processing, Reasoning, and Learning (SPIRAL), Electrical and Computer Engineering Department, Northeastern University, Boston, MA 02115 USA

Mathew Yarossi,

The Movement Neuroscience Laboratory, Department of Physical Therapy Rehabilitation and Movement Science, Northeastern University, Boston, MA 02115 USA.

Tales Imbiriba,

Cognitive Systems Laboratory, Electrical and Computer Engineering Department, Northeastern University, Boston, MA 02115 USA; Center for Signal Processing, Reasoning, and Learning (SPIRAL), Electrical and Computer Engineering Department, Northeastern University, Boston, MA 02115 USA

Dana H. Brooks [Life Senior Member IEEE],

Cognitive Systems Laboratory, Electrical and Computer Engineering Department, Northeastern University, Boston, MA 02115 USA; Center for Signal Processing, Reasoning, and Learning (SPIRAL), Electrical and Computer Engineering Department, Northeastern University, Boston, MA 02115 USA

Eugene Tunik,

The Movement Neuroscience Laboratory, Department of Physical Therapy Rehabilitation and Movement Science, Northeastern University, Boston, MA 02115 USA.

Deniz Erdo mu [Senior Member IEEE]

Cognitive Systems Laboratory, Electrical and Computer Engineering Department, Northeastern University, Boston, MA 02115 USA; Center for Signal Processing, Reasoning, and Learning (SPIRAL), Electrical and Computer Engineering Department, Northeastern University, Boston, MA 02115 USA

Abstract

Transcranial Magnetic Stimulation (TMS) can be used to map cortical motor topography by spatially sampling the sensorimotor cortex while recording Motor Evoked Potentials (MEP) with surface electromyography (EMG). Traditional sampling strategies are time-consuming and inefficient, as they ignore the fact that responsive sites are typically sparse and highly spatially

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see <https://creativecommons.org/licenses/by/4.0/>

Corresponding author: Razieh Faghihpirayesh. raziehfaghih@ece.neu.edu .

correlated. An alternative approach, commonly employed when TMS mapping is used for presurgical planning, is to leverage the expertise of the coil operator to use MEPs elicited by previous stimuli as feedback to decide which loci to stimulate next. In this paper, we propose to automatically infer optimal future stimulus loci using active learning Gaussian Process-based sampling in place of user expertise. We first compare the user-guided (USRG) method to the traditional grid selection method and randomized sampling to verify that the USRG approach has superior performance. We then compare several novel active Gaussian Process (GP) strategies with the USRG approach. Experimental results using real data show that, as expected, the USRG method is superior to the grid and random approach in both time efficiency and MEP map accuracy. We also found that an active warped GP entropy and a GP random-based strategy performed equally as well as, or even better than, the USRG method. These methods were completely automatic, and succeeded in efficiently sampling the regions in which the MEP response variations are largely confined. This work provides the foundation for highly efficient, fully automatized TMS mapping, especially when considered in the context of advances in robotic coil operation.

Keywords

Active learning; entropy; Gaussian process; machine learning; motor cortex; motor evoked potentials; transcranial magnetic stimulation

I. Introduction

TMS has become a valuable noninvasive method to map the motor cortical representation of a specific muscle [1]. The basic procedure is to apply TMS pulses to a variety of locations on the scalp overlying the sensorimotor cortex while recording EMGs from target muscle(s), and quantifying the amplitude of resultant MEPs. The resulting “map” of MEP amplitudes as a function of TMS pulse positions describes an estimate of the topographic organization of the muscle(s) under study. The cortical muscle topographies identified by TMS mapping have shown high correlation to the areas identified by other more expensive and burdensome imaging modalities such as fMRI and PET [2]-[4]. Consequently, TMS mapping has become a popular tool for research studies and clinical applications such as tracking cortical reorganization and neurosurgical planning. The current standard technique for mapping involves stimulating with the TMS coil 3-10 times at each site on a regular grid, most commonly consisting of 25-100 points at a 0.5-1 cm spacing. These points are typically identified on the scalp by physically drawing on a swimmer’s cap or using computer-guided navigation systems [5], [6]. Given the large number of points, and a recommended interstimulus time on the order of several seconds, mapping procedures may take as long as one hour to complete. In practice, much of this time is spent recording null responses, since the majority of the stimulation points typically lie outside of the excitable area for the muscle of interest [7]. This time scale prohibits measurement of transient cortical stimulation or learning-induced changes and is not well tolerated by clinical populations. Previous approaches to improving the efficiency of mapping have included stimulating in a spiral pattern by moving radially out from a pre-determined locus or “hotspot” to reduce null stimulation points [8], or by replacing the grid with a cross-like pattern [7] aligned to the

precentral gyrus. Recently, a pseudo-random walk sampling pattern combined with interpolation of excitability among the stimulation sites was reported to produce statistically similar outcomes to grid-based approaches with comparatively fewer stimuli [9]-[11].

When TMS mapping is used for surgical planning, the need for greater efficiency has led to an operator dependent (USRG) approach, in which stimulus locations are chosen by the TMS coil operator based on real time feedback of observed MEP responses [12]. The operator is free to increase or decrease stimulation point density at their discretion. This typically reduces the number of null stimulations while simultaneously increasing stimulation density in responsive areas. A number of studies on the use of TMS for preoperative planning using this approach have reported favorable outcomes [12]-[14], including high correlation to “gold standard” maps produced using intracranial electrical stimulation [3], [15].

The USRG approach could be categorized as a form of implicit active learning (a field of machine learning) [16], [17], since the coil operator utilizes an implicit objective function built on experience and expertise to select stimulation loci. The goal of this work is to replace the TMS operator’s expertise with an automated algorithm based on *statistical* active learning. This kind of automated approach will not only result in more objective, time efficient, and repeatable sampling; it will also be particularly useful for TMS mapping that utilizes new advances in robotic coil positioning [18], as it allows the development of fully autonomous, efficient mapping. Figure 1 illustrates such a framework, in which TMS operator expertise in the USRG mapping approach can be replaced with robotic coil control guided by active learning.

The topic of active sample selection has been investigated in many fields including applications to optimal sensor placement [19]-[21], active GP sampling [22], and weather forecasting [23]. Ideally, given an optimality criterion, the goal is to select a limited number of samples that “best” represents the phenomenon of interest (e.g., here the field of MEP responses). One approach to this problem statement is to employ greedy on-line strategies where at each step the algorithm searches for the next sample location that leads to the highest one-step gain in the chosen optimality criterion. One statistical approach that can be used to model a spatial phenomena is Gaussian Processes (GPs), an approach for probabilistic regression in reproducing kernel Hilbert spaces that allows us to explicitly model uncertainty about estimated function values. GPs provide a well-defined yet flexible stochastic framework which can be exploited to compute and optimize the desired criteria [24]. Crucially, GPs provide both a spatial interpolation and an estimate of model uncertainty as a function of space [20]. These two quantities are associated with the Gaussian representation of the spatial field and allow the derivation of analytical expressions for conditional entropy that are typically needed for information theoretic criteria. In recent studies, several optimal point selection strategies have been presented in conjunction with GPs [19], [20], [24]. In these works, the authors considered both entropy-based and mutual information (MI)-based criteria, providing greedy, near-optimal strategies. These strategies succeeded in reducing the GP’s spatial uncertainty, because the variances provided by the GP depend exclusively on the *location* of points, while completely neglecting available information regarding the field’s *amplitude* variations. This implies that even when the

changes in the field's amplitude is confined to a specific subregion, these methodologies will lead to a relatively even sampling of the entire space (as determined in detail by the selection of the GP kernel, which governs its spatial covariance structure) [16]. In problems such as TMS mapping, where the Region of Interest (ROI) is completely or largely confined to a subregion of the space, this tendency towards evenly-spaced sampling reduces resolution in the region where the activity is concentrated. A good criterion for automated TMS mapping should balance exploration and exploitation to discover the ROI and concentrate most of the samples in that region.

We previously reported an initial formulation of an alternative GP-based information-theoretic strategy relying on amplitude uncertainty instead of spatial uncertainty [16]. This was accomplished by modeling the mean of the GP as a GP itself, and considering information-based criteria which led to selecting samples that reduced uncertainty across MEP amplitude values. One drawback of this formulation was that it neglects spatial uncertainty, causing it to get stuck in areas with high MEP amplitudes while ignoring unexplored regions of the space—in effect favoring exploitation over exploration. Inspiration for alternative approaches can be found in the use of active sampling in other domains. Specifically, Gunter *et al.* proposed an active sampling strategy as a tool for Bayesian Quadrature integration which leveraged the fact that the GP regression is being used to approximate likelihood functions, and, thus, must be non-negative [25]. To match this observation, the authors considered GPs with a square-root warping function (see (5)) which has several desirable properties, including guaranteed non-negativity for the observation model, and a covariance function which naturally balances exploration and exploitation. Another alternative sampling algorithm might be considered where the balance between exploration and exploitation depends exclusively on the shape of the GP mean modeled as a probability density function (pdf). For instance, Imbiriba *et al.* proposed a GP-based sampling algorithm developed in the context of particle filtering, where samples were drawn from a GP mean function that was normalized in order to make it a viable pdf [26]. This strategy presents some features that fit well with TMS mapping. First, it naturally concentrates most samples in the ROI. Second, it also leads to random exploration of the space in which the compromise with exploitation will be automatically controlled by the concentration of the field's amplitude.

In this paper we investigate the GP-based active learning strategies discussed above for the automated TMS mapping problem. We report on two studies. First, we compared the performance of experimental (traditional) sampling approaches (regular grid sampling (GRID), random sampling (RAND), and USRG) acquired from 5 individuals, and second, we compared the performance of computational GP-based strategies for the reconstruction of the experimentally acquired maps with respect to the time efficiency and accuracy of the maps. We hypothesized that a computational GP-based strategy can be used to achieve at least similar accuracy to an expert user-guided sampling approach with equal or greater efficiency.

II. Gaussian Process for TMS Mapping

GPs can be viewed as interpolators capable of providing Gaussian distributions for every point in the space. This stochastic formulation provides important features that are relevant here, in particular that GPs provide model uncertainty measures as a function of location. This quantified model uncertainty can be used to aid in trading off between exploration and exploitation. In addition, having a probabilistic model enables the use of methods based on information theoretic criteria. The GP formulation defines stochastic models directly in the functional space by assuming Gaussian priors for functions $\psi \in \mathcal{H}$, $\psi | \mathbf{x} \sim \mathcal{N}(0, \kappa(\mathbf{x}, \mathbf{x}))$, where \mathcal{H} is a functional Hilbert space, κ is a covariance function or kernel, and $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d$ are the inputs of the system of interest. We assume that we observe an ‘‘output’’ \mathbf{y} corresponding to every \mathbf{x} . Then, given a set of N input-output pairs $\{\mathbf{x}_k, \mathbf{y}_k\}_{k \in \mathcal{A}}$, where $\mathbf{y} \in \mathbb{R}$ and \mathcal{A} is an index set such that $|\mathcal{A}| = N$, GP regression aims at providing a predictive distribution for a new sample \mathbf{x}_ℓ conditioned on the training set. This can be accomplished by assuming a Gaussian model relating inputs and outputs:

$$y_k = \psi(\mathbf{x}_k) + \eta_k \quad (1)$$

where $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$. The predictive distribution [27] for a new input \mathbf{x}_ℓ can be obtained as:

$$\psi | y_{\mathcal{A}}, \mathbf{X}_{\mathcal{A}}, \mathbf{x}_\ell \sim \mathcal{N}(\hat{\psi}_{\ell | \mathcal{A}}, s_{\ell | \mathcal{A}}^2) \quad (2)$$

with $\hat{\psi}_{\ell | \mathcal{A}} = \kappa_{\mathcal{A}\ell}^T (\mathbf{K}_{\mathcal{A}\mathcal{A}} + \sigma_\eta^2 \mathbf{I})^{-1} \mathbf{y}_{\mathcal{A}}$, and $s_{\ell | \mathcal{A}}^2 = \kappa_{\ell\ell} - \kappa_{\mathcal{A}\ell}^T (\mathbf{K}_{\mathcal{A}\mathcal{A}} + \sigma_\eta^2 \mathbf{I})^{-1} \kappa_{\mathcal{A}\ell}$, where $\kappa_{\mathcal{A}\ell} = \kappa(\mathbf{X}_{\mathcal{A}}, \mathbf{x}_\ell)$, $\kappa_{\ell\ell} = \kappa(\mathbf{x}_\ell, \mathbf{x}_\ell)$, and $\mathbf{X}_{\mathcal{A}} = [\mathbf{x}_{\mathcal{A}_1}, \dots, \mathbf{x}_{\mathcal{A}_N}]^T$. For a detailed derivation see [27]. The Bayesian framework also provides strategies to estimate free parameters, such as the kernel parameters $\boldsymbol{\theta}$ and the noise power σ_η^2 . We consider the exponentiated-quadratic kernel with $\boldsymbol{\theta} \in \mathbb{R}^2$ containing the scaling and kernel bandwidth parameters, with maximal likelihood kernel parameter selection [27] that minimizes (3):

$$-2 \ln f_Y(y_{\mathcal{A}} | X_{\mathcal{A}}, \sigma_\eta^2, \boldsymbol{\theta}) = y_{\mathcal{A}}^T \mathbf{K}_{\mathcal{A}\mathcal{A}}^{-1} y_{\mathcal{A}} + \ln | (2\pi) \mathbf{K}_{\mathcal{A}\mathcal{A}} | \quad (3)$$

with respect to $(\sigma_\eta^2, \boldsymbol{\theta})$.

GP regression has been successfully used in a wide variety of fields including regression and classification [27], detection [28], [29], unmixing [30], and Bayesian optimization [31]. One limitation of GP is related to the Gaussian assumption on the posterior distribution, which makes standard GP models unfit or inaccurate when the observation distribution is non-Gaussian. This is the case for TMS mapping where measured MEP amplitudes are inherently non-negative quantities. To circumvent this issue for our setting, a mapping function can be applied to Gaussian random variables to enforce non-negativity, resulting in non-Gaussian random variables. These procedures with non-linear mappings are known as Warped GPs (WGPs) [25], [32], which we describe in detail next.

A. Warped Gaussian Process

To guarantee non-negativity of the functions being modeled and at the same time exploit the GP framework, a common strategy is to consider a warping function $g: \mathcal{Y} \subset \mathbb{R}_+ \rightarrow \mathcal{Z} \subset \mathbb{R}$, $Y \mapsto Z$, where Y is a non-negative random variable. Let Z and Y be two multivariate random variables, whose realizations are z and y , related through a mapping function g as $Z = g(Y)$. The densities of Z and Y are related [33] according to (4).

$$f_Y(y) = f_Z(g(y)) \left| J_Y^{-1} \right| \quad (4)$$

where $J_Y = \frac{\partial(z_1, \dots, z_N)}{\partial(y_1, \dots, y_N)}$ is the Jacobian matrix. This allows exchanging learning in the original space for learning in the latent warped space \mathcal{Z} . Thus, using (4), fitting a GP in the latent space reduces to maximizing the GP's log marginal likelihood function [32]. Different warping functions have been considered in the literature to cope with the requirements of different applications [25], [32].

In the context of TMS mapping, warping functions must guarantee non-negativity of the estimated field without imposing hard upward limits. Thus, in this work, we selected the square-root warping function given by:

$$z = g(y) = \sqrt{2(y)}, \quad y = g^{-1}(z) = \frac{1}{2}z^2 \quad (5)$$

III. Active GP Sample Selection

A general problem formulation for active sample selection consists of selecting a set of input-output pairs $\mathcal{A} := \{x_\ell, y_\ell\}_{\ell=1}^{n_{\mathcal{A}}}$, with cardinality $|\mathcal{A}| = n_{\mathcal{A}}$, such that the loci $x_\ell \in \mathcal{R}$ where the measurements y take place are selected following some optimality criterion. In many scenarios where the ROI \mathcal{R} is discretized, i.e., $\ell \in \mathcal{V}$ with \mathcal{V} an index set of all discretized locations in \mathcal{R} , the problem can be stated as optimally selecting the subset \mathcal{A} such that $\ell \in \mathcal{V}$ and $n_{\mathcal{A}} \ll |\mathcal{V}| = n_{\mathcal{V}}$. As mentioned, ideally, given an optimal criterion, the goal is to select $n_{\mathcal{A}}$ samples that “best” represent the ROI. Next, we discuss the main types of information theoretic-based active sample selection methods that will be analyzed in this paper.

A. Standard Information Theoretic-Based Sampling

Information based criteria generally focus on selecting samples that lead to the highest decrease in the uncertainty (entropy) or that are the most informative about unsensed locations (MI) [20]. Since global approaches based on such criteria lead to NP-complete problems [20], [34], greedy iterative strategies are often employed such that, at each iteration, the algorithm will select the locus that leads to the maximum decrease in entropy or maximum increase in MI. Thus, considering \mathcal{V} as defined, and letting \mathcal{A} denote the index set of the loci already selected at some arbitrary iteration, then active selection selects the

index ℓ that leads to the maximum gain according to the selected criteria at each iteration, that is:

$$\ell^* = \arg \max_{\ell \in \mathcal{V} \setminus \mathcal{A}} \delta_{\ell}^{\text{Criterion}} \quad (6)$$

where $\delta_{\ell}^{\text{Criterion}}$ is the adopted criterion quantity and $\mathcal{V} \setminus \mathcal{A}$ is the set of all indexes in \mathcal{V} that are not in \mathcal{A} .

When considering entropy, the goal of GP-based greedy algorithms is to, iteratively, select a locus that results in the largest conditional entropy given by (7) among all of available measurement locations $\ell \in \mathcal{V} \setminus \mathcal{A}$, conditioned on measured loci $\mathbf{y}_{\mathcal{A}}$:

$$H(\psi_{\ell} | \mathbf{y}_{\mathcal{A}}) = - \int p(\boldsymbol{\psi}_{\ell}, \mathbf{y}_{\mathcal{A}}) \log p(\boldsymbol{\psi}_{\ell} | \mathbf{y}_{\mathcal{A}}) d\boldsymbol{\psi}_{\ell} d\mathbf{y}_{\mathcal{A}} \quad (7)$$

The conditional entropy for Gaussian random variables [35] is well known and given by (8).

$$H(\psi_{\ell} | \mathbf{y}_{\mathcal{A}}) = \frac{1}{2} \log(2\pi s_{\ell | \mathcal{A}}^2) + \frac{1}{2} \quad (8)$$

where $s_{\ell | \mathcal{A}}^2$ is the GP variance, leading to a simple and effective selection strategy. Since $\log(\zeta)$ is monotonically increasing for $\zeta \in \mathbb{R}_+$, problem (6) can be solved for GP entropy by finding the sample index $\ell \in \mathcal{V} \setminus \mathcal{A}$, that maximizes the quantity:

$$\delta_{\ell}^{\text{Entropy}} = s_{\ell | \mathcal{A}}^2 = \kappa_{\ell\ell} - \boldsymbol{\kappa}_{\mathcal{A}\ell}^{\text{T}} \mathbf{K}_{\mathcal{A}\mathcal{A}}^{-1} \boldsymbol{\kappa}_{\mathcal{A}\ell}. \quad (9)$$

A known issue with the entropy criterion is that it has the tendency to select loci along the edges of the sample space. This can be explained by the fact that the entropy grid approach aims at selecting the $\boldsymbol{\psi}_{\ell}$ with the largest variance $s_{\ell | \mathcal{A}}^2$, and these uncertainties are known to be larger on the edges of the sampling space [36].

An alternative approach is based on the MI of random variables within the set \mathcal{A} and in $\mathcal{V} \setminus \mathcal{A}$. This strategy leads to an optimization criterion that searches for the subset of locations that most significantly reduces the uncertainty about the estimates in the rest of the space [20]. Problem (6) can be solved for GP MI by finding the sample index $\ell \in \mathcal{V} \setminus \mathcal{A}$ that maximizes the quantity (for more detail see [20]):

$$\delta_{\ell}^{\text{MI}} = \frac{\kappa_{\ell\ell} - \boldsymbol{\kappa}_{\mathcal{A}\ell}^{\text{T}} \mathbf{K}_{\mathcal{A}\mathcal{A}}^{-1} \boldsymbol{\kappa}_{\mathcal{A}\ell}}{\kappa_{\ell\ell} - \boldsymbol{\kappa}_{\bar{\mathcal{A}}\ell}^{\text{T}} \mathbf{K}_{\bar{\mathcal{A}}\bar{\mathcal{A}}}^{-1} \boldsymbol{\kappa}_{\bar{\mathcal{A}}\ell}} \quad (10)$$

where $\bar{\mathcal{A}} = \mathcal{V} \setminus (\mathcal{A} \cup \ell)$, and $\mathbf{K}_{\mathcal{Q}\mathcal{Q}}$ is the kernel matrix for all loci \mathbf{x}_{ℓ} with $\ell \in \mathcal{Q}$.

MI-based strategies, different from entropy, tend to find loci that are most informative about the unstimulated locations, thus avoiding sampling the edges [20]. One drawback of MI active sampling strategies is the large computational complexity of the resulting greedy algorithms, since at each iteration the computation of (10) requires matrix inverses whose sizes are proportional to all points available ($|\mathcal{A}|$).

Previously, we showed that although MI can lead to faster initial convergence because it does not initially focus on sampling the edges, for the TMS mapping problem, the two methods perform very similarly in practice as the number of samples increases, providing loci that are almost uniformly placed on the scalp over the motor cortex [16]. For this reason, we only consider entropy-based strategies in this paper.

B. Amplitude-Based Entropy Sampling

To overcome the limitation that naive entropy-based active sampling strategies tend to lead to uniformly distributing sampling loci, in our previous study we proposed an alternative approach that takes into account the variations of the function ψ , leading to a sampling strategy that seemed intuitively to more closely match the USRG strategy [16]. We used the GP estimate itself (i.e., the GP mean) $\hat{\psi}$ to construct an iterative strategy where GP estimation and sample selection were performed sequentially. Thus, given an index set \mathcal{A} with respective sample pairs $\{X_{\mathcal{A}}, y_{\mathcal{A}}\}$ the GP estimate $\hat{\psi}_p$ for all $p \in \mathcal{T}$ can be obtained by taking the mean of (2). Assuming a zero-mean Gaussian prior for $\hat{\psi}_p$ with covariance $\kappa(\cdot, \cdot)$, we have

$$\hat{\psi}_p \sim \mathcal{N}(0, \tilde{\kappa}_{pp}) \quad (11)$$

with $\tilde{\kappa}_{pp} \triangleq \kappa(\hat{\psi}_p, \hat{\psi}_p)$.

The MEPs at selected indices $\hat{\Psi}_{\mathcal{A}}$ are distributed

$$\hat{\Psi}_{\mathcal{A}} \sim \mathcal{N}(0, \tilde{\mathbf{K}}_{\mathcal{A}\mathcal{A}}) \quad (12)$$

with $\tilde{\mathbf{K}}_{\mathcal{A}\mathcal{A}} \triangleq \kappa(\hat{\Psi}_{\mathcal{A}}, \hat{\Psi}_{\mathcal{A}})$. Now, if we consider a sample $\hat{\psi}_\ell$, $\ell \in \mathcal{T} \setminus \mathcal{A}$, the joint distribution of $\hat{\psi}_\ell$ and $\hat{\Psi}_{\mathcal{A}}$ is given by

$$\begin{pmatrix} \hat{\Psi}_{\mathcal{A}} \\ \hat{\psi}_\ell \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \tilde{\mathbf{K}}_{\mathcal{A}\mathcal{A}} & \tilde{\kappa}_{\mathcal{A}\ell} \\ \tilde{\kappa}_{\mathcal{A}\ell}^\top & \tilde{\kappa}_{\ell\ell} \end{pmatrix}\right), \quad (13)$$

with $\tilde{\kappa}_{\mathcal{A}\ell} \triangleq [\kappa(\hat{\psi}_{\mathcal{A}_1}, \hat{\psi}_\ell), \dots, \kappa(\hat{\psi}_{\mathcal{A}_{n_{\mathcal{A}}}}, \hat{\psi}_\ell)]^\top$. Using the identity in [27] we have

$$\hat{\psi}_\ell | \hat{\Psi}_{\mathcal{A}} \sim \mathcal{N}\left(\mu_{\hat{\psi}_\ell | \mathcal{A}}, \sigma_{\hat{\psi}_\ell | \mathcal{A}}^2\right) \text{ with}$$

$$\mu_{\hat{\psi}_{\ell} | \mathcal{A}} = \tilde{\kappa}_{\mathcal{A}\ell}^{\text{T}} \tilde{\mathbf{K}}_{\mathcal{A}\mathcal{A}}^{-1} \tilde{\boldsymbol{\Psi}}_{\mathcal{A}}, \quad \sigma_{\hat{\psi}_{\ell} | \mathcal{A}}^2 = \tilde{\kappa}_{\ell\ell} - \tilde{\kappa}_{\mathcal{A}\ell}^{\text{T}} \tilde{\mathbf{K}}_{\mathcal{A}\mathcal{A}}^{-1} \tilde{\kappa}_{\mathcal{A}\ell}. \quad (14)$$

Adopting the entropy-based criterion for sample selection with this amplitude-based approach leads to a sample choice criterion: $\delta_{\ell}^{\text{Criterion}} \triangleq \delta_{\psi, \ell}^{\text{Entropy}}$ as:

$$\delta_{\psi, \ell}^{\text{Entropy}} = \sigma_{\hat{\psi}_{\ell} | \mathcal{A}}^2 = \tilde{\kappa}_{\ell\ell} - \tilde{\kappa}_{\mathcal{A}\ell}^{\text{T}} \tilde{\mathbf{K}}_{\mathcal{A}\mathcal{A}}^{-1} \tilde{\kappa}_{\mathcal{A}\ell}. \quad (15)$$

C. Entropy With Warped GPs Sampling

To ensure non-negativity, Warped GPs can be leveraged as discussed in Section II-A where the invertible mapping $g: \mathcal{Y} \rightarrow \mathcal{Z}$ between the MEP space \mathcal{Y} and a latent space \mathcal{Z} was presented. Thus, equation (5) leads to $Z = g(Y) = \sqrt{2Y}$. Now, assuming a Gaussian prior on $Z: Z | \mathbf{x} \sim \mathcal{N}(0, \kappa(\mathbf{x}, \mathbf{x}))$ the predictive distribution for a new location \mathbf{x}_{ℓ} can be written as:

$Z | z_{\mathcal{A}}, \mathbf{X}_{\mathcal{A}}, \mathbf{x}_{\ell} \sim \mathcal{N}\left(\mu_{z_{\ell} | \mathcal{A}}, \sigma_{z_{\ell} | \mathcal{A}}^2\right)$ with

$$\mu_{z_{\ell} | \mathcal{A}} = \kappa_{\mathcal{A}\ell}^{\text{T}} \mathbf{K}_{\mathcal{A}\mathcal{A}}^{-1} z_{\mathcal{A}}, \quad \sigma_{z_{\ell} | \mathcal{A}}^2 = \kappa_{\ell\ell} - \kappa_{\mathcal{A}\ell}^{\text{T}} \mathbf{K}_{\mathcal{A}\mathcal{A}}^{-1} \kappa_{\mathcal{A}\ell}. \quad (16)$$

Since linear transformations of GPs are also GPs, considering a local linearization of the inverse mapping $Y = \frac{1}{2}Z^2$ leads to a GP over Y given a GP defined over Z [25]. Specifically, considering the first order Taylor expansion of g^{-1} , we have

$$y \simeq g^{-1}(z_0) + \frac{\partial g^{-1}(z)}{\partial z} \Big|_{z=z_0} (z - z_0).$$

Algorithm 1 Active GP Entropy-Based Sampling Algorithm

Input : $X_{\mathcal{A}_0}, y_{\mathcal{A}_0}, \mathcal{V}, X_{\mathcal{V}}, N$, method

Output: \mathcal{A}

- 1 Obtain estimated field $\widehat{\Psi}_{\mathcal{V}}$ by taking the mean of (2) (for GPE and GPE _{μ}) or (18) (for WGPE);
- 2 Set $n = n_0$ and $\mathcal{A} = \mathcal{A}_0$;
- 3 **while** $n \leq N$ **do**
 - for** $\ell \in \mathcal{V} \setminus \mathcal{A}$ **do**
 - Compute $\delta_{\Psi, \ell}$ using (9) or (15) or (20) based on method;
 - end**
 - 4 $\ell^* = \arg \max_{\ell} \delta_{\widehat{\Psi}, \ell}$;
 - Get MEP sample y_{ℓ^*} at location x_{ℓ^*} ;
 - $\mathcal{A} = \mathcal{A} \cup \{\ell^*\}$;
 - $\mathcal{V} = \mathcal{V} - \{\ell^*\}$;
 - Obtain estimated field $\widehat{\Psi}_{\mathcal{V}}$ by taking the mean of (2) or (18);
 - Increment n ;
- 13 **end**
- 14 **return** \mathcal{A} ;

One interesting characteristic of the mapping g is that the derivative $\frac{\partial g^{-1}(z)}{\partial z} = z$. Thus, choosing $z_0 = \mu_{z_\ell | \mathcal{A}}$ leads to:

$$y \simeq \frac{1}{2} \mu_{z_\ell | \mathcal{A}}^2 + \mu_{z_\ell | \mathcal{A}} (z - \mu_{z_\ell | \mathcal{A}}) \quad (17)$$

and using the approximation $z \simeq \mu_{z_\ell | \mathcal{A}}$, we have:

$$Y | y_{\mathcal{A}}, X_{\mathcal{A}}, x_\ell \sim \mathcal{N}(\mu_{y_\ell | \mathcal{A}}, \sigma_{y_\ell | \mathcal{A}}^2) \quad (18)$$

with

$$\mu_{y_\ell | \mathcal{A}} = \frac{1}{2} \mu_{z_\ell | \mathcal{A}}^2, \quad \sigma_{y_\ell | \mathcal{A}}^2 = \mu_{z_\ell | \mathcal{A}}^2 \sigma_{z_\ell | \mathcal{A}}^2. \quad (19)$$

Consequently, the entropy criterion leads to solving problem (6) using (19):

$$\delta_{Z,\ell}^{WGP} = \sigma_{y_{\ell}|\mathcal{A}}^2 = \mu_{z_{\ell}|\mathcal{A}}^2 \sigma_{z_{\ell}|\mathcal{A}}^2. \quad (20)$$

One important consequence of the selected mapping g and the linearization of its inverse is the fact that the variance in (19) and, consequently, the quantity (20), are functions of the GP's variance and mean. This implies that this approach balances exploration and exploitation naturally by selecting loci where the model has high uncertainty (large $\sigma_{z_{\ell}|\mathcal{A}}^2$) and high amplitude (large $\mu_{z_{\ell}|\mathcal{A}}$).

D. Entropy Based Sampling Algorithm

Algorithm 1 presents the proposed active GP entropy-based sampling strategies. It is designed to work with a set of initial loci $X_{\mathcal{A}_0}$ and their respective MEPs $y_{\mathcal{A}_0}$, a set of selectable indices \mathcal{V} with associated loci $X_{\mathcal{V}}$, the final cardinality N , and finally the chosen entropy-based method. In line 7, a new index ℓ^* is selected based on a chosen method given by (9), or (15), or (20). We refer to these methods as: standard GP entropy (**GPE**), Amplitude-based GP entropy (**GPE _{μ}**), and Warped GP entropy (**WGPE**) respectively. The algorithm follows an iterative sequence interchanging between computing $\delta_{\psi,\ell}$ for all $\ell \in \mathcal{V} \setminus \mathcal{A}$ (lines 4–6), finding the optimal index (line 7), updating the sets and MEP field estimation (lines 9–11). When the desired cardinality (N) is achieved the algorithm returns the set of selected (suggested) stimuli loci and the corresponding estimated MEPs.

To have a baseline for the purpose of comparison we also implement a naive uniform grid sampling (**UGRID**) strategy by considering a uniform grid of N points over the space at each iteration.

IV. Adaptive GP Random Sampling

An entirely different possible sampling strategy for TMS mapping would consist of obtaining sample locations by randomly selecting locations according to an appropriate distribution $p(\mathbf{x})$ defined over \mathcal{X} . A naive approach would be to uniformly sample over the entire motor cortex. A more principled approach, however, would exploit the knowledge provided by iteratively observing the MEP values acquired by previous samples to update an estimate of this distribution and, thus, gradually concentrate future samples in the area where previous MEPs had larger amplitudes (namely, ROI).

Imbiriba *et al.* proposed a GP-based sampling methodology in the context of particle filtering where, at each iteration, the GP's mean function was used to approximate the posterior distributions of particle filters and in the resampling process [26]. To adopt this strategy to the problem at hand we need to choose an appropriate model for the probability density function (PDF) to update. We note that a function $p(\mathbf{x})$ is a valid PDF if it obeys two basic properties: 1) non-negativity and 2) it integrates to one. Non-negativity can be achieved here by leveraging Warped GPs as discussed in Section II-A while the integration-

to-one requirement can be achieved by properly normalizing the WGP mean function. That is, if we let $\mu(\mathbf{x})$ be the mean function of a non-negative WGP, then

$$\hat{p}(\mathbf{x}) = \mu(\mathbf{x}) / \int \mu(\mathbf{x}') d\mathbf{x}' \quad (21)$$

is a valid PDF.

Algorithm 2 presents the Gaussian Process Random Sampling (**GPRS**) algorithm. It receives as inputs an initial set of loci and measurements $\{\mathbf{X}_0, \mathbf{y}_0\}$ with cardinality n_0 , the final number of points N and the initial distribution $\hat{p}_{n_0}(\mathbf{x})$. Then, an iterative procedure is

performed by randomly sampling a new location $\mathbf{x}_\ell \sim \hat{p}_\ell(\mathbf{x})$, measuring a new MEP y_ℓ at \mathbf{x}_ℓ and updating the distribution by using all available pairs \mathbf{X}, \mathbf{y} . The update performed in line 6 involves updating the WGP and normalizing it according to (21). The sampling procedure at line 3 can be performed using any strategy aimed at sampling arbitrary distributions such as Accept-Reject, Importance Sampling, etc [37]. In this work, we opted for the Accept-Reject method due to its simplicity. In order to have a baseline for purpose of comparison we also implement a naive random sampling strategy by considering $p(\mathbf{x})$ as a fixed uniform distribution in the area of interest (e.g. over the motor cortex). We refer to this approach as Uniform Random Sampling (**URAND**).

Algorithm 2 Adaptive GP Random Sampling Algorithm

Input : $\mathbf{X}_0, \mathbf{y}_0, n_0, n_{\mathcal{A}}, \hat{p}_{n_0}(\mathbf{x})$
Output: \mathbf{X}, \mathbf{y}
1 $\mathbf{X} = \mathbf{X}_0$;
2 **for** $\ell = n_0, \dots, N$ **do**
 | Sample \mathbf{x}_ℓ from $\hat{p}_\ell(\mathbf{x})$;
 | Get MEP y_ℓ at loci \mathbf{x}_ℓ ;
3 | Make $\mathbf{X} = [\mathbf{X}^\top, \mathbf{x}_\ell^\top]^\top$ and $\mathbf{y} = [\mathbf{y}^\top, y_\ell]^\top$;
 | Update $\hat{p}_\ell(\mathbf{x})$ using the pair $\{\mathbf{X}, \mathbf{y}\}$;
7 **end**
8 **return** \mathbf{X}, \mathbf{y} ;

An interesting difference between the main two sampling strategies proposed in this paper (Active GP Entropy-based and Adaptive GP Random Sampling) is that the latter does not require a grid and incurs a much lower computational cost, since it does not need to evaluate the quantity $\delta_\ell^{\text{Criterion}}$ for all points in a grid. On the other hand, its results are, obviously, variable from one run to the next due to the random draws from the density.

V. Experimental Setup

In this section we present our experimental setup. First, we discuss the non-automatic TMS mapping strategies used to collect real data. Then, we discuss how we used this collected data to build a realistic data set with accessible “ground truth”. Finally, we present the metrics used for assessing the performance of the different methodologies.

A. TMS Mapping Procedure

All protocols were conducted in conformance with the Declaration of Helsinki and were approved by the Institutional Review Board of Rutgers Biomedical Health Sciences, Newark, NJ. Five healthy, right handed subjects (3 males & 2 females) aged 22 to 40 participated following written informed consent. Participants were free from neurological and musculoskeletal deficits, and systemic disease. Exclusion criteria included contraindications for TMS, including metallic or electronic implants in the head, pregnancy, and history of epilepsy. Subjects were seated with their arm, hand, and fingers comfortably secured in a brace to limit motion. Surface electromyographic activity (EMG, Delsys Trigno, 1kHz) was recorded from the first dorsal interosseous (FDI) of the right hand. To record the location of TMS (Magstim, Rapid²) stimuli, each subject’s head was coregistered to a canonical high-resolution anatomical MRI for frameless neuronavigation (Advanced Neuro Technology). The TMS coil (Magstim, 70 mm double coil) was held tangential to the scalp, with the handle posterior 45° off the sagittal plane [7], [38], [39]. MEPs were quantified as the peak-to-peak (pp) amplitude of the FDI EMG signal in a window from 10 to 40 ms following the TMS pulse. The cortical hotspot was found by sampling the hemisphere until the locus with the largest MEP was located [8], [40]. Resting motor threshold (RMT) was determined at this location as the minimum intensity required to elicit MEPs $> 50 \mu V$ in the FDI muscle on 3 of 6 consecutive trials [38], [41]. All TMS measures were taken at rest and background EMG was monitored to ensure that muscles remained relaxed. All mapping was performed with the stimulation intensity set to 110% of RMT [42]. Each mapping included 294 stimulations, delivered at an interstimulus interval of $4.2 \pm .25$ seconds, necessitating ~20 minutes to complete. Mapping of the right FDI representation on the left sensorimotor cortex using a GRID, RAND, or USRG-based spatial sampling approach were performed in a single session with 5-10 minute breaks between mappings. Order of the mapping was randomized between subjects. Details of each method are as follows: 1) GRID: A 7×7 grid (49 points, 36 cm^2) with 1 cm spacing between grid sites was centered at the FDI hotspot in the tangential plane to the scalp. To place the grid on the scalp, head model data was exported from the neuronavigation software into Matlab (Mathworks). Grid points were translated inward towards the scalp along a vector intersecting the center point of the standard head model in order to project the grid directly onto the scalp. The grid was then imported into the neuronavigation software and superimposed virtually on the head model. Each of these grid sites was stimulated 6 times. 2) RAND: The random sampling included the 4 corners of the grid, the FDI hotspot, and 289 uniform randomly distributed sites (294 total) over the same $6 \times 6 \text{ cm}$ area. These sites were again projected onto the scalp using the same procedure described for GRID. One stimulus was delivered to each site in the random set. 3) USRG: Only the 4 corners of the grid and the site signifying the FDI hotspot were projected onto the scalp. At the beginning of the experiment, stimulation was applied to each

of these sites once (5 stimulations). The remaining 289 stimulations were chosen at the discretion of the TMS operator based on real-time feedback of MEP amplitude. The intention of the operator was to maximize the information obtained per stimulus by increasing the density of points in what they judged at the time to be excitable and border regions while placing few points in what they judged to be null-response areas [12], [13].

B. TMS Dataset for Computational Experiments

To test the computational methods, we created an accessible “ground truth” (GT) function from the *combination* of three experimental maps (GRID, RAND, USRG). We then used this function as a surrogate to determine the MEP amplitude obtained from any arbitrary sampling location chosen by an algorithm. This is crucial for our computational methods that use the real-time feedback of MEP amplitude to select future loci. To do so, two interpolation steps were used. In the first step cubic interpolation was used to interpolate all available points to a 3.75 mm resolution grid as has been previously described by our group and others [4], [38], [43]. We then used this interpolated data as a training data for WGP model (over GP, to ensure non-negativity), calculated the posterior, and computed the predictive posterior distribution (see section II) on any point of interest \mathbf{x} . Thus, for any of the methods tested, once a sampling location is determined, its associated MEP amplitude can be obtained from this predictive distribution.

For the Active GP Entropy-based Sampling methods that need a set of selectable loci (for greedy approach), we used a grid with 2 mm resolution in 2D space. This resolution seemed good based on real world TMS mapping resolution and map liability. We discuss more about this resolution in the section VII. For Adaptive Random Sampling methods, as stated before, we do not require a grid and methods should choose any arbitrary location in the space.

C. Assessment of Accuracy of the Sampling Methods

To assess the accuracy of the methods, we created a testing data including a dense grid of 0.5 mm resolution loci in 2D space and their associated MEP amplitudes (GT values). Reconstruction error was defined as the normalized mean squared error (NMSE) between GT MEP values and predicted MEP values from selected loci by each method for testing data. Specifically, we defined NMSE as: $\text{NMSE}(\boldsymbol{\psi}, \hat{\boldsymbol{\psi}}) = (\|\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}\|_2^2) / (\|\boldsymbol{\psi}\|_2^2)$ where $\boldsymbol{\psi}$ is a vector containing the GT MEPs and $\hat{\boldsymbol{\psi}}$ is a corresponding vector of MEPs obtained using the sample points chosen by a particular method. In addition to NMSE, we computed several common map features as outcome measures: map volume, map area, and center of gravity in both the rostral-caudal plane (denoted COGx1) and in the medial-lateral plane (COGx2). Map volume and map area were calculated using double trapezoidal integration of the interpolated maps [44]. COGx1 and COGx2 were computed using standard equations [42].

VI. Results

Mapping procedures were well tolerated by all 5 subjects and no adverse effects of the stimulation were reported. Stimulation loci selected by each experimental method (GRID, RAND, USRG) are shown for subject 3 in Figure 2. Map features (area, volume, COGx1, and COGx2) determined for the maximum number of sampled loci are shown for the GT

and each collected map in the left-most columns (Experimental Methods) in Table I. Group Mean (GM), defined as the average across subjects, MapArea and MapVol were observed to be smaller for GRID sampling, than for RAND or USRG sampling (see last row of Table I). To illustrate how the error in the acquired experimental maps depends on the method as more samples are acquired, the evolution of NMSE between GT and each experimental map across the duration of mapping is shown for all subjects in the first row of Figure 4. Lower GM NMSE was observed for USRG and RAND when compared with the GRID approach (GM NMSE: 0.36 ± 0.11 , 0.68 ± 0.17 , 0.99 ± 0.00 respectively) after 49 simulations. The GRID approach resulted in less reduction of NMSE with added stimuli compared to RAND or USRG. USRG outperformed RAND (GM NMSE: 0.33 ± 0.10 , 0.60 ± 0.18 respectively) following 100 stimuli for all 5 subjects, and for some subjects increased performance was seen with fewer stimuli. RAND and USRG became more similar in terms of NMSE with added stimuli in all subjects; however USRG provided a better fit to the GT in every case.

For all computational methods, an initial uniform grid set \mathcal{A}_0 with cardinality $|\mathcal{A}_0| = 36$ was selected. We selected a total of 256 stimuli with each of the computational methods. Figure 3 presents details of the loci of the stimuli selected by each method for subject 3. To evaluate the extent to which each algorithm samples the excitable region of the cortex the percentage of sampled stimuli with a MEPS greater than $50\mu V$ (the threshold used to determine RMT, see Sec. V-A) were determined. Averaged across participants, the methods that placed a high percentage of new stimuli in the ROI (excitable region) were GPE_μ ($67\pm 16\%$), GPRS ($79\pm 12\%$), and WGPE ($78\pm 13\%$), while UGRID ($12\pm 4\%$), URAND ($14\pm 3\%$), and GPE ($12\pm 3\%$) spread stimuli more uniformly over the motor cortex. This resulted in the GPE_μ , GPRS, and WGPE approaches being qualitatively more similar to the USRG approach ($40\pm 5\%$ stimuli in ROI) both in terms of behavior as a function of number of stimuli as well as the final sample map appearance and progression. Figure 4 (second row) shows the evolution of NMSE between the GT and the maps from each of the computational methods as a function of increasing number of stimuli. Monte Carlo simulations were performed for methods with randomness (URAND and GPRS) and for those methods we show the mean (solid color) and standard deviation (STD) (transparent shade) in the figure. Lower GM NMSE after 100 simulations was observed for USRG, GPRS, and WGPE (0.33 ± 0.10 , 0.24 ± 0.10 , 0.14 ± 0.07 respectively) when compared with GPE_μ , GPE, URAND, and UGRID (0.82 ± 0.07 , 0.48 ± 0.20 , 0.58 ± 0.19 , 0.52 ± 0.27 respectively). The best performing methods, WGPE and GPRS, achieved lower NMSE than the operator-dependent USRG method at GM number of stimuli: 65.25 ± 13.93 and 90.25 ± 10.33 respectively. In 4 of 5 participants lower NMSE was consistently maintained with increased sampling (Figure. 4).

As an additional comparison, we defined the number of stimuli needed to produce an NMSE below 0.1 as a reasonable standard at which to compare. Among all methods WGPE, GPRS, and USRG methods required the fewest number of stimuli (GM: 122.4 ± 27.5 , 140.4 ± 24.5 , and 184.0 ± 36.96 respectively) to reach this threshold.

Evaluating results in terms of map features rather than NMSE, in Figure 5 we report results for those features for all subjects and all methods as a function of increasing number of

stimuli, where the straight red lines in the plots show the GT value for each feature. Visual inspection of these figures suggests that some methods (e.g. UGRID and GPE_{μ}) experienced high variability as the number of stimuli increased. USRG, WGPE, and GPRS typically were within 10% of the true value at full cardinality for all metrics across the subjects. UGRID, URAND, GPE, and GPE_{μ} all performed less accurately. For example, GPE_{μ} only reaches to within 15% and 28% of the true value across the features in subjects 3 and 5 respectively.

In addition to the results for the experimental methods, Table I also reports on results for the map features determined at full cardinality for the computational methods. In terms of the coefficient of variation (CV,¹ relative standard deviation) for each feature, COGx1 and CoGx2 (CV = 0.038, and 0.0168 respectively) had the least variability across the different computational approaches in comparison to MapVol and MapArea (CV = 0.114, 0.0824 respectively). Across individuals MapArea results was consistent with USRG for the WGPE and GPRS methods, slightly larger for UGRID and GPE, and smaller for URAND and GPE_{μ} . MapVol was observed to be consistent with USRG for WGPE, GPRS, UGRID, and GPE approaches while URAND and GPE_{μ} approaches were characterized by smaller MapVol (see last row of Table I).

VII. Discussion

As a first step towards highly efficient fully automated TMS mapping, in this study we described and compared several computational methods, motivated by the USRG approach (in which future stimulation loci are selected utilizing online feedback of resultant amplitude), for the selection of future stimuli. In terms of performance, among the experimental applied methods, the USRG sampling pattern was found to more rapidly reduce NMSE (relative to the GT) with added stimuli compared to RAND or GRID approaches, and resulted in the lowest NMSE at full cardinality for all subjects. The large reduction in NMSE in the first 50 stimuli is indicative of the active selection process made by the operator to maximize information using as few stimuli as possible. Using a pseudo-random walk pattern, van de Ruit and colleagues reported the production of maps with high correlation to the 100 stimuli full map in as few as 50 stimuli [9] (GM 63 stimuli). However, the use of at least 80 stimuli was recommended in order to generate a random sampling distribution that adequately covered the response area and maps that were less susceptible to test-retest variability. Our results showing greater efficiency (reduction of NMSE with fewer stimuli) of the USRG approach compared to a random approach confirms the suggestion in [9]. However, our random procedure was different than the one used by van de Ruit *et al.* [9]. In this study, 184 stimuli, representing 20% of the number of stimuli used to define the GT, were needed to meet our accuracy requirements, possibly indicating better performance than shown in van de Ruit *et al.* [9]. However, comparisons between our study and van de Ruit *et al.* [9] of the total number of stimuli needed to generate an accurate map are difficult, given differences in the total number of stimuli used to construct the GT and differences in the metric used to evaluate accuracy. Interestingly, results showed relatively small improvements in NMSE with increased stimuli in the GRID condition. While this may seem

¹ $CV = \sigma/\mu$, where σ is the standard deviation and μ is the mean of population.

surprising, a recent investigation reported that only two stimuli per grid site were needed to create maps of equal validity to those obtained when five stimuli per site were used [10], consistent with our results.

The USRG approach can be characterized as a balance between exploitation of current knowledge about the excitable area and exploration of areas where excitability is uncertain. The relative success or failure of our computational approaches appears to be based on how each approach handled this exploration vs exploitation tradeoff. Among all computational methods tested, the GPRS and WGPE approaches had similar performance to the USRG method although they handle exploration-exploitation trade-off efficiently in two very distinct ways. For GPRS, exploration is introduced by the random nature of the algorithm while exploitation is introduced as the evolution of the GP mean modifies the probability of sampling regions of the space as new samples are obtained. For WGPE, exploration-exploitation is a consequence of the quantity (20) that scales the latent GP variance with the square of its mean, thereby increasing exploitation as the mean function evolves over time. These features underlie our confidence that WGPE and GPRS are the best strategies among the algorithms discussed in this study. On the other hand, the UGRID, URAND, and GPE methods all tend to place samples evenly in the space, neglecting exploitation.

As stated before, all the computational methods except URAND and GPRS are grid-based, meaning the algorithm selects the next stimulus from a finite grid (selectable set) without replacement. In our experiments, we used a finite 2D grid with 2mm resolution as our selectable set, based on the limitations of the coil tracking software. However, we noticed that these approaches were influenced by the preallocated grid. This mostly affected methods based on both GP spatial uncertainty and MEP amplitude like WGPE and GPE_{μ} . Indeed GPE_{μ} was heavily affected by this factor, because it focuses only on the variations of MEP values (exploitation) and completely neglects spatial exploration. This is illustrated by comparing the GPE_{μ} results of this paper, in which a 2 mm selectable grid was used, to those of our previous paper [16] in which a 3 mm selectable grid was used. With a sparser grid, the algorithm is forced to select loci that are farther apart from each other (at the expense of losing definition of sharp peaks), whereas with the denser grid used here it got stuck exploiting small regions. This behaviour can be seen in Figure 3 where the loci selected by GPE_{μ} are overly concentrated in part of the ROI while ignoring unexplored regions.

A. Limitations

A number of limitations may have impacted the findings of this investigation. A canonical brain was used for neuronavigated TMS instead of a reconstruction from the participants MRI. This may have impacted the TMS operator in the USRG condition, who may have used the subject-specific anatomy to guide coil position early during mapping. All experiments were conducted by an experienced TMS operator. It is unknown if the same result would be found if the individual selecting the stimulation sites was naïve to TMS mapping. This should be the subject of future investigation, as well as the inter-operator reliability. Only a single muscle was mapped. It is unknown whether multi-muscle mapping would be improved by a user-guided approach. To maximize such an approach, the TMS operator would need to be given feedback from each muscle collected. In this case, an

automated or even random approach may better capture multiple representations at once. The coil orientation was fixed at 45° to midline (approximately perpendicular to the central sulcus) and was not adjusted for sulcus alignment on each individual stimulation. The method used to determine resting motor threshold was a less common alternative to the more common Rossini-Rothwell method [45] (which uses 5 of 10 stimulation greater than 50 μV) as a threshold. Given maps for each participant were conducted with the same coil orientation and intensity in the same session, we do not believe that the coil orientation or method to determine RMT had an impact on the active sampling results presented here. Incorporation of the coil orientation and intensity is an achievable goal for improvement to our algorithm. In the future we intend update our algorithms to operate directly on the induced electric field for individual head models thereby incorporating coil orientation and stimulus intensity as well as three dimensional individual anatomy.

VIII. Conclusion

In this paper, we leveraged different GP-based active learning strategies to study possibilities for automated TMS mapping, and compared these methods with traditional strategies. Warped GP entropy-based sampling and GP random sampling methods were found to be the best candidates for a fully automated replacement for operator-dependent mapping. We note that our study of different active sampling methods has potential applications to a wide set of spatial prediction problems outside the scope of TMS mapping.

Acknowledgments

This project was supported by NIH grants R01NS085122 (ET), 2R01HD058301 (ET), R01DC009834 (DE), R01NS111744 (DE), and F31NS092268 (MY), and NSF grants IIS-1715858 (DE), IIS-1149570 (DE), and CNS-1544895 (DE), and CBET-1804550 (ET, DE, DB).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Rutgers Biomedical Health Sciences, Newark, NJ, under Approval No. Pro0120080216.

References

- [1]. Thickbroom G, Byrnes M, and Mastaglia F, "Methodology and application of TMS mapping," *Electroencephalogr. Clin. Neurophysiol. Suppl.*, vol. 51, p. 48, 1. 1999. [PubMed: 10590935]
- [2]. Classen Jet al., "Multimodal output mapping of human central motor representation on different spatial scales," *J. Physiol.*, vol. 512, no. 1, pp. 163–179, 10. 1998. [PubMed: 9729626]
- [3]. Tarapore PEet al., "Preoperative multimodal motor mapping: A comparison of magnetoencephalography imaging, navigated transcranial magnetic stimulation, and direct cortical stimulation: Clinical article," *J. Neurosurg.*, vol. 117, no. 2, pp. 354–362, 8. 2012. [PubMed: 22702484]
- [4]. Weiss Cet al., "Mapping the hand, foot and face representations in the primary motor cortex—Retest reliability of neuronavigated TMS versus functional MRI," *NeuroImage*, vol. 66, pp. 531–542, 2. 2013. [PubMed: 23116812]
- [5]. Lüdemann-Podubecká J and Nowak DA, "Mapping cortical hand motor representation using TMS: A method to assess brain plasticity and a surrogate marker for recovery of function after stroke?" *Neurosci. Biobehav. Rev.*, vol. 69, pp. 239–251, 10. 2016. [PubMed: 27435238]
- [6]. Gunduz MEet al., "Motor cortex reorganization in limb amputation: A systematic review of TMS motor mapping studies," *Frontiers Neurosci.*, vol. 14, p. 314, 4. 2020.

- [7]. Littmann AE, McHenry CL, and Shields RK, “Variability of motor cortical excitability using a novel mapping procedure,” *J. Neurosci. Methods*, vol. 214, no. 2, pp. 137–143, 4. 2013. [PubMed: 23357026]
- [8]. Koski L, Mernar TJ, and Dobkin BH, “Immediate and long-term changes in corticomotor output in response to rehabilitation: Correlation with functional improvements in chronic stroke,” *Neurorehabil. Neural Repair*, vol. 18, no. 4, pp. 230–249, 12. 2004. [PubMed: 15537994]
- [9]. van de Ruit M, Perenboom MJL, and Grey MJ, “TMS brain mapping in less than two minutes,” *Brain Stimulation*, vol. 8, no. 2, pp. 231–239, 3. 2015. [PubMed: 25556004]
- [10]. Cavaleri R, Schabrun SM, and Chipchase LS, “The reliability and validity of rapid tms mapping,” *Brain Stimulation*, vol. 11, no. 6, pp. 1291–1295, 2018. [PubMed: 30025980]
- [11]. Jonker ZDet al., “TMS motor mapping: Comparing the absolute reliability of digital reconstruction methods to the golden standard,” *Brain Stimulation*, vol. 12, no. 2, pp. 309–313, 3. 2019. [PubMed: 30497885]
- [12]. Niskanen E, Julkunen P, Säisänen L, Vanninen R, Karjalainen P, and Könönen M, “Group-level variations in motor representation areas of the thenar and anterior tibial muscles: Navigated TMS study,” *Hum. Brain Mapping*, vol. 31, no. 8, pp. 1272–1280, 2010.
- [13]. Krieg SMet al., “Protocol for motor and language mapping by navigated TMS in patients and healthy volunteers; workshop report,” *Acta Neurochir*, vol. 159, no. 7, pp. 1187–1195, 7. 2017. [PubMed: 28456870]
- [14]. Picht T, Frey D, Thieme S, Kliesch S, and Vajkoczy P, “Presurgical navigated TMS motor cortex mapping improves outcome in glioblastoma surgery: A controlled observational study,” *J. Neuro-Oncol*, vol. 126, no. 3, pp. 535–543, 2. 2016.
- [15]. Picht Tet al., “Preoperative functional mapping for rolandic brain tumor surgery: Comparison of navigated transcranial magnetic stimulation to direct cortical stimulation,” *Neurosurgery*, vol. 69, no. 3, pp. 581–589, 9. 2011. [PubMed: 21430587]
- [16]. Faghihpirayesh R, Imbiriba T, Yarossi M, Tunik E, Brooks D, and Erdo mu D “Motor cortex mapping using active Gaussian processes,” in *Proc. 13th ACM Int. Conf. Pervas. Technol. Rel. Assistive Environ.*, 6. 2020, pp. 1–7.
- [17]. Sourati J, Akcakaya M, Leen TK, Erdogmus D, and Dy JG, “Asymptotic analysis of objectives based on Fisher information in active learning,” *J. Mach. Learn. Res*, vol. 18, no. 1, pp. 1123–1163, 2017.
- [18]. Finke M, Fadini T, Kantelhardt S, Giese A, Matthaus L, and Schweikard A, “Brain-mapping using robotized TMS,” in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 8. 2008, pp. 3929–3932.
- [19]. Guenther J and Sawodny O, “Optimal sensor placement based on Gaussian process regression for shared office spaces under various ventilation conditions,” in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, 10. 2019, pp. 2883–2888.
- [20]. Krause A, Singh A, and Guestrin C, “Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies,” *J. Mach. Learn. Res*, vol. 9, no. 2, pp. 235–284, 2008.
- [21]. Sourati J, Akcakaya M, Erdogmus D, Leen TK, and Dy JG, “A probabilistic active learning algorithm based on Fisher information ratio,” *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 40, no. 8, pp. 2023–2029, 8. 2018. [PubMed: 28858784]
- [22]. Huber MF, “Recursive Gaussian process: On-line regression and learning,” *Pattern Recognit. Lett*, vol. 45, pp. 85–91, 8. 2014.
- [23]. Comboul M, Emile-Geay J, Hakim GJ, and Evans MN, “Paleoclimate sampling as a sensor placement problem,” *J. Climate*, vol. 28, no. 19, pp. 7717–7740, 10. 2015.
- [24]. Schulz E, Speekenbrik M, and Krause A, “A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions,” *J. Math. Psychol*, vol. 85, pp. 1–16, 8. 2018.
- [25]. Gunter T, Osborne MA, Garnett R, Hennig P, and Roberts SJ, “Sampling for inference in probabilistic models with fast Bayesian quadrature,” in *Proc. NIPS*, 2014, pp. 2789–2797.
- [26]. Imbiriba T and Closas P, “Enhancing particle filtering using Gaussian processes,” in *Proc. IEEE 23rd Int. Conf. Inf. Fusion (FUSION)*, 7. 2020, pp. 1–7.

- [27]. Williams CKI and Rasmussen CE, Gaussian Processes for Machine Learning, vol. 2, no. 3. Cambridge, MA, USA: MIT Press, 2006.
- [28]. Imbiriba T, Bermudez JCM, Tourneret J-Y, and Richard C, "Detection of nonlinear mixtures using Gaussian processes: Application to hyperspectral imaging," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 5 2014, pp. 7949–7953.
- [29]. Imbiriba T, LaMountain G, Wu P, Erdogmus D, and Closas P, "Change detection and Gaussian process inference in piecewise stationary environments under noisy inputs," in Proc. IEEE 8th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP), 12. 2019, pp. 530–534.
- [30]. Imbiriba T, Bermudez JCM, Richard C, and Tourneret J-Y, "Nonparametric detection of nonlinearly mixed pixels and endmember estimation in hyperspectral images," IEEE Trans. Image Process, vol. 25, no. 3, pp. 1136–1151, 3. 2016. [PubMed: 26685243]
- [31]. Frazier PI, "A tutorial on Bayesian optimization," 2018, arXiv:1807.02811. [Online]. Available: <http://arxiv.org/abs/1807.02811>
- [32]. Snelson E, Ghahramani Z, and Rasmussen CE, "Warped Gaussian processes," in Proc. NIPS, 2004, pp. 337–344.
- [33]. Papoulis A and Pillai SU, Probability, Random Variables, and Stochastic Processes. New York, NY, USA: McGraw-Hill, 2002.
- [34]. Ko CW, Lee J, and Queyranne M, "An exact algorithm for Maxent sampling," Oper. Res, vol. 43, no. 4, pp. 684–691, 1995.
- [35]. Cover TM and Thomas JA, Elements of Information Theory. Hoboken, NJ, USA: Wiley, 2012.
- [36]. James G, Witten D, Hastie T, and Tibshirani R, An Introduction to Statistical Learning, vol. 112. New York, NY, USA: Springer, 2013.
- [37]. Robert C and Casella G, Monte Carlo Statistical Methods. New York, NY, USA: Springer, 2013.
- [38]. Yarossi Met al., "The association between reorganization of bilateral M1 topography and function in response to early intensive hand focused upper limb rehabilitation following stroke is dependent on ipsilesional corticospinal tract integrity," Frontiers Neurol., vol. 10, p. 258, 3. 2019.
- [39]. Souza VH, Vieira TM, Peres ASC, Garcia MAC, Vargas CD, and Baffa O, "Effect of TMS coil orientation on the spatial distribution of motor evoked potentials in an intrinsic hand muscle," Biomed. Eng./Biomedizinische Technik, vol. 63, no. 6, pp. 635–645, 11. 2018.
- [40]. Sollmann Net al., "Inter- and intraobserver variability in motor mapping of the hotspot for the abductor pollicis brevis muscle," BMC Neurosci., vol. 14, no. 1, p. 94, 2013. [PubMed: 24006839]
- [41]. Trompetto C, Assini A, Buccolieri A, Marchese R, and Abbruzzese G, "Motor recovery following stroke: A transcranial magnetic stimulation study," Clin. Neurophysiol, vol. 111, no. 10, pp. 1860–1867, 10. 2000. [PubMed: 11018503]
- [42]. Ngomo S, Leonard G, Moffet H, and Mercier C, "Comparison of transcranial magnetic stimulation measures obtained at rest and under active conditions and their reliability," J. Neurosci. Methods, vol. 205, no. 1, pp. 65–71, 3. 2012. [PubMed: 22227444]
- [43]. Borghetti Det al., "TMS mapping: A model based on spline interpolation," Brain Res. Bull, vol. 77, nos. 2–3, pp. 143–148, 2008. [PubMed: 18588952]
- [44]. Yarossi M, Adamovich S, and Tunik E, "Sensorimotor cortex reorganization in subacute and chronic stroke: A neuronavigated TMS study," in Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 8. 2014, pp. 5788–5791.
- [45]. Rossini Pet al., "Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: Basic principles and procedures for routine clinical application. Report of an IFCN committee," Electroencephalogr. Clin. Neurophysiol, vol. 91, no. 2, pp. 79–92, 8. 1994. [PubMed: 7519144]

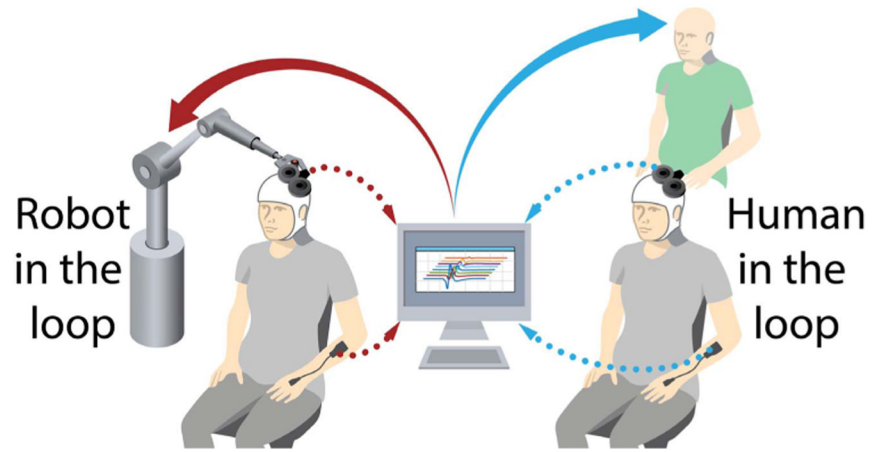


Fig. 1. Illustration of a framework where operator expertise (right) for selection of stimuli in the user-guided mapping approach is replaced with machine learning and robotics(left). Here, a machine-learning-driven robot is shown to illustrate autonomous mapping.

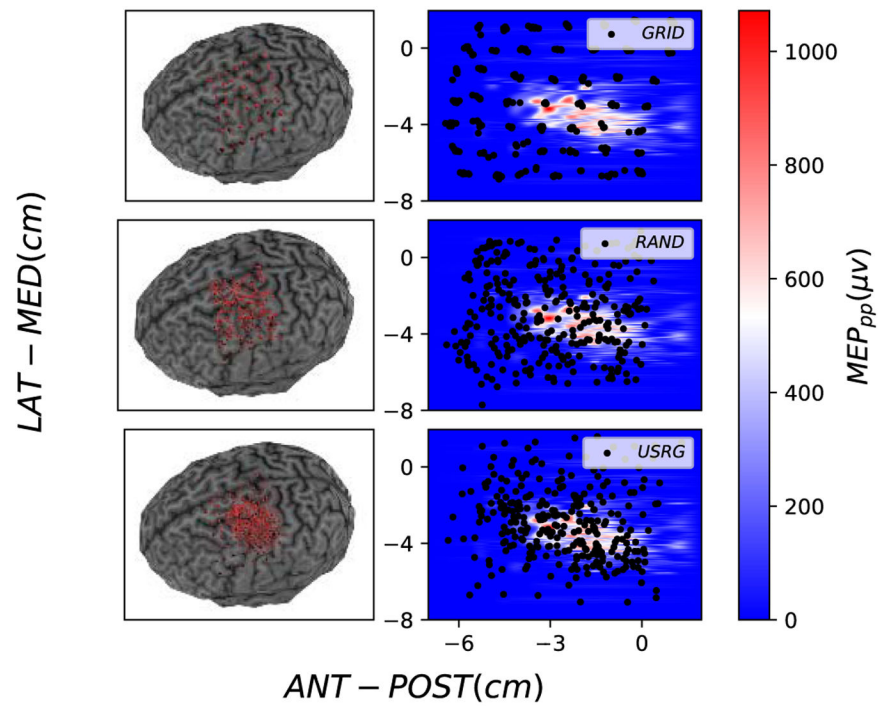


Fig. 2. Stimulation loci (Left: red, right: black circles) for GRID (top), RAND (middle) and USRG (bottom) mappings shown on the **Left:** 3D curvilinear canonical brain used for neuronavigation and **Right:** Heat maps indicating interpolated 2D map representation of MEP amplitude in μV for subject 3. The color bar represents the MEP_{pp} and axes display the spatial coordinate system used by the neuronavigation software which places the origin at the intersection of the fiducial markers used for co-registration (roughly at the center of the brain).

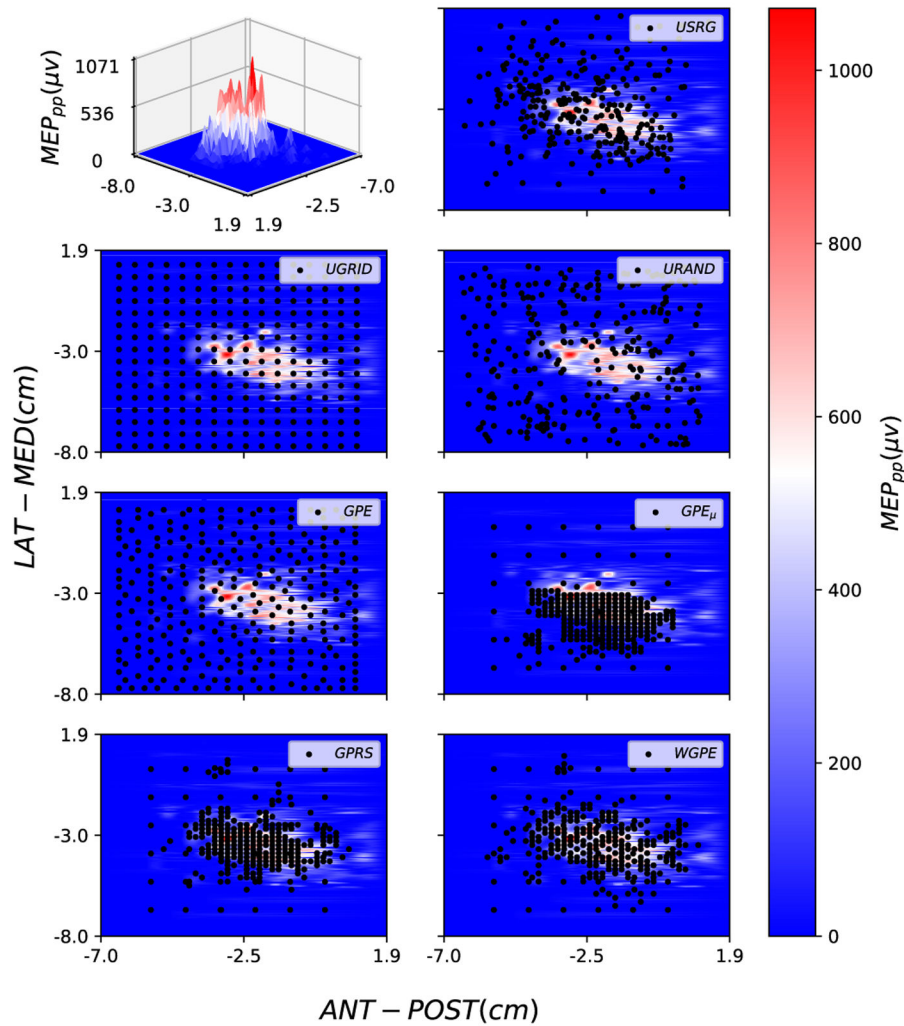


Fig. 3. Using Warped GP model, a surface is fitted on the real TMS data to create a representation of MEP map for subject 3. This map can be viewed in either a 3D (top left) or 2D (others) map. The color bar represents the MEP_{pp} and Black circles indicate the selected stimulation loci by each method (specified in top right of each subplot).

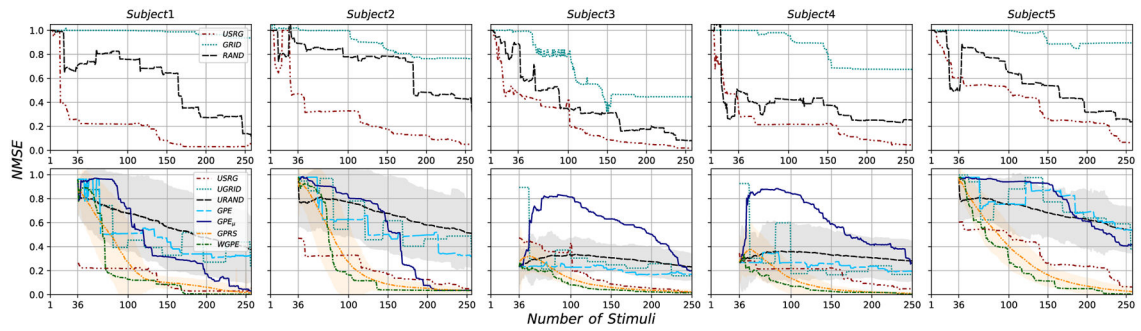


Fig. 4. NMSE between GT and predicted MEP values for testing data as the number of stimuli are added by each experimental method (**First row**) and computational method (**Second row**) for all subjects. NMSE mean (line) and Standard deviations (transparent shade) are presented for URAND and GPRS methods for which Monte Carlo procedures were used to understand randomness in stimuli selection.

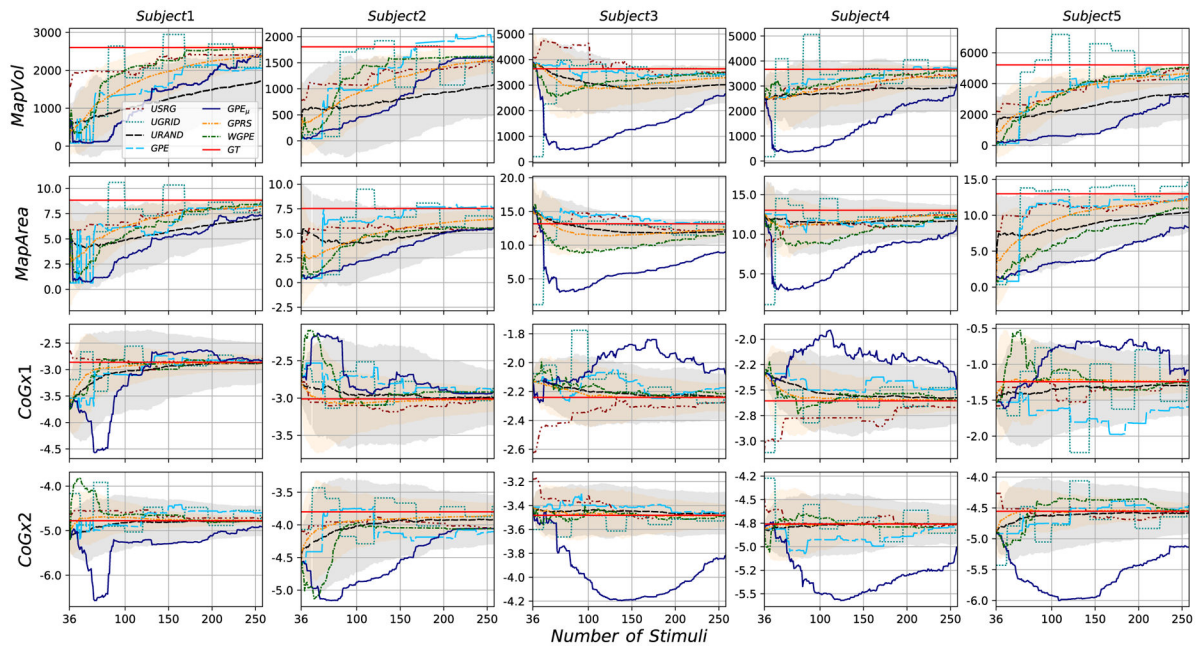


Fig. 5. Map features evolution for all subjects and all methods. The ground truth's values are represented by the red constant lines.

TABLE I

Map Features of the Final Map Generated With Every Method for All Subjects. The Ground Truth Values Are Shown in the Column Labeled GT. The Columns Show Both Experimental (Left) and Computational (Right) Results. Units Are Given for S1 and Are Identical for Each Feature for the Other Subjects. The Group Mean and Standard Deviations for Volume and Area Are Shown at the Bottom

	MAP Features	GT	Experimental Methods					Computational Methods				
			USRG	GRID	RAND	UGRID	URAND	GPE	GPE _μ	GPRS	WGPE	
S1	Volume (μV/cm ²)	2603.48	2450.64	195.80	2294.27	2190.25	1751.59±774.48	2251.15	2550.75	2560.36±91.19	2573.36	
	Area (cm ²)	8.83	7.78	1.07	7.03	8.08	7.12±1.77	8.68	8.16	8.34±0.57	8.41	
S2	CoGx1 (cm)	-2.86	-2.77	-2.99	-2.72	-2.80	-2.86±0.27	-2.96	-2.86	-2.86±0.05	-2.87	
	CoGx2 (cm)	-4.78	-4.72	-4.62	-4.70	-4.70	-4.76±0.27	-4.64	-4.81	-4.77±0.05	-4.79	
	Volume	1806.05	1541.48	625.29	1369.29	1655.72	1101.68±702.86	1899.03	1618.74	1657.39±82.48	1644.27	
	Area	7.53	5.49	3.69	7.10	7.45	5.63±2.22	7.55	5.66	5.99±0.67	5.86	
S3	CoGx1	-3.01	-2.98	-2.24	-3.04	-3.10	-2.98±0.38	-2.90	-2.93	-2.98±0.08	-2.95	
	CoGx2	-3.79	-4.05	-4.10	-3.84	-3.92	-3.90±0.37	-4.00	-4.03	-3.95±0.17	-4.00	
	Volume	3641.94	3518.31	1881.54	3405.88	3417.25	3093.78±636.23	3456.38	3080.61	3529.35±133.33	3504.24	
	Area	13.17	12.52	8.75	12.37	12.18	12.08±1.37	12.99	10.12	12.26±0.88	11.86	
S4	CoGx1	-2.24	-2.28	-2.28	-2.33	-2.25	-2.24±0.13	-2.20	-2.19	-2.23±0.06	-2.20	
	CoGx2	-3.48	-3.46	-3.36	-3.40	-3.45	-3.48±0.14	-3.45	-3.69	-3.51±0.07	-3.48	
	Volume	3671.18	3442.27	1376.30	3021.00	3378.83	2966.97±889.25	3791.60	3554.99	3623.41±82.16	3611.56	
	Area	13.03	12.48	6.68	12.25	12.31	11.53±1.98	11.79	12.13	12.47±0.67	12.24	
S5	CoGx1	-2.61	-2.68	-2.42	-2.61	-2.61	-2.60±0.24	-2.51	-2.51	-2.59±0.04	-2.60	
	CoGx2	-4.76	-4.74	-4.33	-4.72	-4.82	-4.75±0.23	-4.75	-4.83	-4.75±0.04	-4.76	
	Volume	5211.60	4998.68	422.23	3640.00	4661.77	3345.79±1656.68	4456.19	3749.20	5004.62±227.48	5148.00	
	Area	13.00	12.62	1.92	11.14	12.71	10.57±2.60	12.68	8.75	11.47±1.00	12.24	
Group Mean	CoGx1	-1.24	-1.20	-2.12	-1.17	-1.12	-1.24±0.39	-1.45	-0.98	-1.24±0.08	-1.25	
	CoGx2	-4.55	-4.60	-4.02	-4.57	-4.67	-4.58±0.27	-4.60	-4.98	-4.55±0.10	-4.56	
Volume	-	3190.28±1158.19	900.23±2.42	2746.09±825.80	3060.76±1051.04	3170.87±957.32	2910.86±767.56	3275.03±1122.95	3296.29±1168.18			

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

MAP Features	GT	Experimental Methods					Computational Methods				
		USRG	GRID	RAND	UGRID	URAND	GPE	GPE _μ	GPRS	WGPE	
Area	-	10.18±2.98	4.42±2.89	9.98±2.42	10.54±2.28	9.38±2.55	10.73±2.20	8.96±2.14	10.10±2.54	10.12±2.57	