

Towards tailoring non-invasive brain stimulation using real-time fMRI and Bayesian optimization

Romy Lorenz^{*†}, Ricardo P Monti^{*†}, Adam Hampshire[†], Yury Koush[‡], Christoforos Anagnostopoulos[†], Aldo A Faisal[†], David Sharp[†], Giovanni Montana^{†§}, Robert Leech[†] and Ines R Violante[†]

[†]Imperial College London, UK, [‡]EPFL, Switzerland, [§]King's College London, UK

^{*} These authors contributed equally to this work.

Abstract—Non-invasive brain stimulation, such as transcranial alternating current stimulation (tACS) provides a powerful tool to directly modulate brain oscillations that mediate complex cognitive processes. While the body of evidence about the effect of tACS on behavioral and cognitive performance is constantly growing, those studies fail to address the importance of subject-specific stimulation protocols. With this study here, we set the foundation to combine tACS with a recently presented framework that utilizes real-time fMRI and Bayesian optimization in order to identify the most optimal tACS protocol for a given individual. While Bayesian optimization is particularly relevant to such a scenario, its success depends on two fundamental choices: the choice of covariance kernel for the Gaussian process prior as well as the choice of acquisition function that guides the search. Using empirical (functional neuroimaging) as well as simulation data, we identified the squared exponential kernel and the upper confidence bound acquisition function to work best for our problem. These results will be used to inform our upcoming real-time experiments.

Keywords—Bayesian optimization, fMRI, non-invasive brain stimulation, transcranial alternating current stimulation

I. INTRODUCTION

Studies involving non-invasive brain stimulation have reported remarkable changes in cognitive and behavioral performance [1]. Among those techniques, transcranial alternating current stimulation (tACS) is particularly promising as it directly allows for the modulation of physiologically relevant brain oscillations that subservise cognitive operations [2]. The effects of tACS are highly dependent on the frequency and phase of stimulation [3]. The conventional tACS approach involves defining the frequency and phase of stimulation ad hoc and testing them on a cohort of subjects. However, this approach exhibits two limitations: (1) the brain networks targeted by the stimulation cannot be verified without simultaneous fMRI; (2) those stimulation parameters may vary across subjects due to difference in anatomy or heterogeneity in disease. Yet, there is a combinatorial explosion in the biologically plausible range of stimulation frequencies (1-100 Hz) and phases (0-359°), resulting in up to thousands of possibilities. Identifying the optimal stimulation protocol for a given individual is like “finding a needle in a haystack”. Therefore, using conventional methodology makes tailoring tACS to an individual highly unfeasible. To address this fundamental challenge in the clinical use of tACS, we aim to combine tACS with a recently presented framework that behavior real-time functional magnetic resonance imaging (fMRI) and Bayesian optimization: The *Automatic Neuroscientist* [4]. Employing

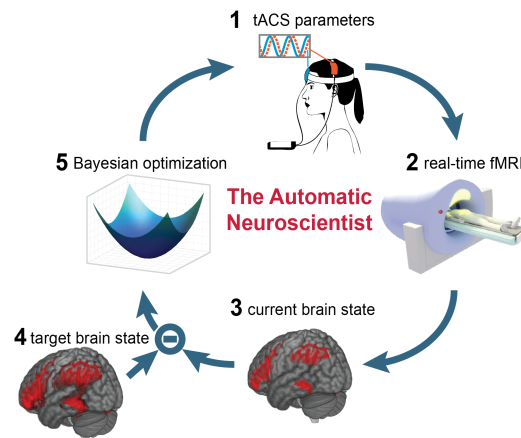


Fig. 1. High-level overview of the *Automatic Neuroscientist* combined with tACS. (1) The experiment starts with applying a random combination of tACS parameters to the subject in the scanner. (2) Whole-brain functional images are acquired and analyzed in real-time in response to the block of stimulation. (3) Information about the current brain state is extracted and (4) compared to the pre-defined target brain state. This result is then fed into the Bayesian optimization approach. (5) Based on this, the algorithm chooses a new combination of tACS parameters that optimizes for the target brain state. This closed-loop cycle continues until the optimal tACS parameters are found.

such a framework allows us to start with a target brain state and find a set of tACS parameters (frequency and phase of stimulation) for each individual that maximally activates it (Fig. 1).

Bayesian optimization is particularly relevant to such a problem as it provides a powerful strategy for finding maxima of objective functions that are expensive to evaluate and might contain noisy measurements (both criteria are met with functional neuroimaging data). Moreover, Bayesian optimization is suited in situations where we do not have an analytical expression of the objective function nor can make formal statements regarding its properties [5], [6]. However, the implementation of Bayesian optimization requires several fundamental choices that are paramount to the success of the technique; such as (1) the choice of covariance function for the Gaussian process (GP) prior and (2) the choice of acquisition function which determines the manner in which the space of parameters is explored.

The aim of this study is to objectively compare the performance of a variety of distinct kernel functions (resulting in distinct GP priors) and acquisitions functions. Offline fMRI

data from eight healthy volunteers receiving blocks of non-invasive tACS stimulation over left frontoparietal brain regions as well as simulation data were employed. The results obtained will be used to propose a combination of kernel and acquisition functions for upcoming real-time experiments.

II. METHODS

The objective of this work is to obtain a better understanding of the relationship between various combinations of tACS parameters (i.e., frequency and phase) and each subject’s neural response. In the framework of Bayesian optimization, such a relationship is summarized by a latent objective function which we wish to infer. With this in mind, a small neuroimaging study was conducted where healthy participants received tACS consisting of various combinations of frequency and phase. This data was subsequently employed to perform model selection on the GP covariance function. In addition to this, simulation analyses were carried out to compare two popular acquisition functions.

A. Empirical Data

1) *Subjects and Experimental Design:* Eight healthy volunteers (5 female, 24.75 ± 3.49 years) took part in our study. The study was approved by the Hammersmith Hospital (London, UK) research ethics committee. Each participant performed three runs of a switch task [7] in the MR scanner. Each run consisted of task blocks (36 s) interleaved with rest blocks (24 s). During task blocks, tACS was applied via a pair of MR-compatible conductive rubber electrodes over left frontal (F3) and left parietal (P3) regions (as determined by the International 10-20 EEG system). Both return electrodes were placed on the ipsilateral shoulder. The current was fixed to 1 mA (peak-to-peak). The experiment parameter space was limited to ten different frequencies (0, 1, 5, 8, 12, 16, 20, 26, 40 and 80 Hz) and five different phases (0, 60, 90, 180 and 270°), resulting in 50 different frequency-and-phase combinations (see Fig. 2). For each run and each subject, we tested a subset of 14 identical frequency-and-phase combinations (grey shaded squares in Fig. 2). The order of those frequency-and-phase combinations was random for each run in order to eliminate any order effect of the stimulation on the subject’s neural response. The condition without any stimulation (0 Hz with 0°) was the only condition that was repeated two (n=4) to four (n=4) times within each run.

2) *Target brain state extraction:* Whole-brain coverage images were acquired by a Siemens Verio 3T scanner using an EPI sequence (TR: 1.5s). Data was minimally preprocessed by applying motion-correction, high-pass filtering (100s) and spatial smoothing (5mm FWHM Gaussian kernel). We then extracted 20 regions of interest (ROIs) that have been previously shown to predict trial-by-trial performance in task switching (for ROI definition we created 8mm spheres around the peak coordinates reported in [8]). The extracted timecourses were further cleaned by removing high-frequency noise and large signal spikes using a modified Kalman filter [9]. We then assessed changes in functional connectivity between those regions for blocks of tACS using psychophysiological interaction (PPI) analyses [10] (whilst also including six motion parameters as confounds). For the analyses presented here,

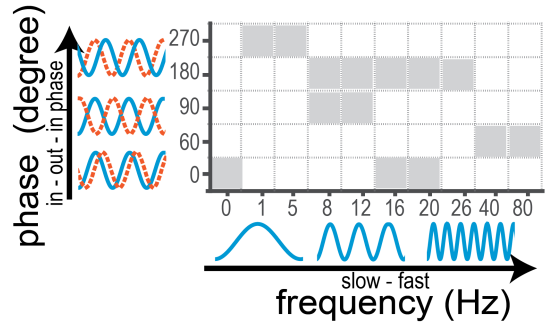


Fig. 2. Exhaustive two-dimensional tACS parameter space with each dimension corresponding to the frequency or phase of non-invasive brain stimulation, respectively. Grey shaded squares are combinations sampled in our (offline) fMRI study.

we selected the two regions that exhibited the highest inter-run robustness across all subjects (assessed using Spearman’s rank correlation coefficient): the left inferior parietal lobule and the posterior cingulate gyrus. In order to assess how likely the pattern of results between those regions could have occurred by chance, we performed non-parametric permutation testing (10,000 permutations) and found the result to be highly significant ($z=3.71$, $p<.001$).

B. Bayesian optimization

The underlying intuition behind the method of Lorenz et al. [4] is that the target brain pattern is a function of experimental conditions. As such, the authors propose to learn the relationship by modelling the observed brain state as a sample from a Gaussian process (GP). This facilitates the use of a Bayesian optimization framework [5], [6] in a closed-loop form where subjects are presented with an experimental condition and real-time fMRI provides instantaneous information about the subject’s brain state. Based on this, we can iteratively update our beliefs about the unknown objective function, captured in the posterior distribution of the GP. While Lorenz et al. [4] used simple perceptual stimuli in their original experiment, our study here involves blocks of non-invasive brain stimulation with different combinations of frequency and phase. This can be considered as a far more challenging problem as only a handful of studies exist investigating the effects of tACS on the blood oxygenation level dependent (BOLD). Beyond that, high inter-subject variability can be expected as outlined in the Introduction.

In contrast to the fMRI setting employed in [4], there is far less prior knowledge regarding the properties of the latent objective function which can be leveraged in this context. Therefore, data was collected over a cohort of eight subjects (see previous section). Using this empirical data we begin by studying the choice of kernel function in Section II-B1. Given a choice of kernel and by carrying out simulation analyses, the effects of various acquisition functions are subsequently assessed in Section II-B2.

1) *Covariance function:* The choice of kernel function in GP regression is fundamental [11]. The kernel directly specifies a measure of similarity across various inputs and therefore determines the generalization properties of the estimated model. Model selection in the context of GPs is often posed

as a parameter estimation task whereby the hyper-parameters for a specific choice of kernel can be learnt [11]. However, comparing the performance of potentially many distinct kernels is a challenging problem. Recently, [12], [13] have proposed a method for automatically searching over the space of kernel structures in a principled manner.

Briefly, the proposed approach proceeds by considering compositional structures (either addition or multiplication) over four *base kernels*; squared exponential (as employed in [4]), periodic, linear or rational quadratic. When comparing various distinct kernel structures the Bayesian Information Criterion (BIC) is employed to score the various models (where type-2 maximum likelihood estimates of parameters are used). Initially, each of the base kernels is proposed. Thereafter, the selected kernel can be extended by either performing addition or multiplication with any base kernel. While the original approach described in [12] proposes kernels for each dimension independently, in this work both dimensions were studied simultaneously (thus two dimensional kernels were proposed at each step). This decision is based on the hypothesis that each dimension (either frequency or phase) would share a similar relationship with the response, therefore leading to the selection of similar (if not the same) kernel.

2) *Acquisition function*: Of equal importance for the success of the Bayesian optimization is the choice of acquisition function. The role of the acquisition function is to iteratively propose the experimental condition to be presented to the subject. It therefore serves to guide the search over all combinations of parameters (in our case frequency and phase) and must implicitly balance exploration with exploitation. In this work we considered the performance of two popular acquisition functions: the expected improvement (EI) and upper confidence bound (GP-UCB) acquisition function [5]¹. The two acquisition functions differ in their trade-off between exploration and exploitation of the experimental parameters space, giving rise to distinct sampling behaviors over time. Informally, the EI acquisition function can be seen as trying to maximize the expected improvement over the current best. While the GP-UCB also favors the selection of points with high mean value (similar to the EI acquisition function), it also favors points with high variance (i.e., regions worth to explore). In certain settings, this will result in a more explorative behavior of the GP-UCB compared to the more “greedy” EI acquisition function. For algorithmic details, please refer to [5].

While the choice of kernel function can be guided by model selection and information theoretic measures, the comparison of multiple acquisition functions requires explicit knowledge of the underlying objective function. As a result, a simulation study was employed to benchmark the performance of the two aforementioned acquisition functions.

A complex and multimodal objective function was proposed for the simulation study (see Fig. 3a). This was motivated by our empirical results, showing multiple optima and minima for each subject (not shown). As the presence of non-neural noise is well documented for fMRI experiments, we also studied how different levels of contrast-to-noise-ratio

(CNR) affected the Bayesian optimization. We repeated our simulation analyses with CNR values ranging from 0.1 and 1.8 (typically reported CNR values in fMRI literature vary between 0.5 and 1.8 [14]). For each CNR value tested, we ran 100 simulations. The maximum number of iterations was set to 100 with the first five (randomly selected) observations serving as burn-in phase. Thereafter, at each iteration, a new frequency-and-phase combination was proposed by maximizing the respective acquisition function and sampled by the Bayesian optimization algorithm in the next iteration.

As a measure of accuracy we computed spatial correlation between the algorithm’s predictions for the whole parameter space and the “ground truth” objective function. Considering our underlying motivation to gain a holistic understanding of the whole parameter space (learning about maxima and minima), spatial correlations seemed most appropriate to capture similarity across the whole space. This procedure was identical to the one reported in [4].

III. RESULTS

A. Covariance kernel selection

In this work we follow [12], [13], and look to compare several distinct kernels via the use of BIC. Furthermore, we also employ a greedy search algorithm whereby we begin with a base kernel and iteratively extend this kernel.

There are two significant differences in the approaches described in this work to those of [12], [13]. The first is our choice of base kernels; in this work the squared exponential, periodic, linear and Matern ($\frac{3}{2}$) were employed as base kernels. The second difference is that two-dimensional kernels were proposed at each iteration. This choice was motivated by an assumption that the relationship between parameter (frequency or phase) and the response shared a similar functional form. Finally, a minor difference between this work and that of [13] is that change-point functions were not considered here.

Following [12], we begin by proposing each of the base kernels and calculating the associated BIC scores. In this first step the squared exponential kernel was selected. We then proceeded to consider compositional structures. Each of these structures was composed by either adding or multiplying the squared exponential kernel with each of the base kernels. For each compositional kernel, the parameters were estimated using the previous parameters as warm starts. Again, BIC scores were employed to score each of the compositional models. At this stage, there was no reduction in BIC scores resulting in the choice of the squared exponential kernel.

B. Acquisition function selection

Simulation analyses were carried out to assess the performance of the EI and GP-UCB acquisition functions. A range of CNR values were employed in order to recreate the properties of fMRI data as well as study the performance of each of the acquisition functions in scenarios with low signal-to-noise ratio. Throughout this simulation the underlying GP model was maintained constant, allowing us to attribute any differences in performance to the acquisition functions.

For each CNR value, 100 simulations were performed. Each simulation allowed the proposed Bayesian optimization

¹ The probability of improvement (PI) acquisition function was not included as it is typically seen to be over exploitative.

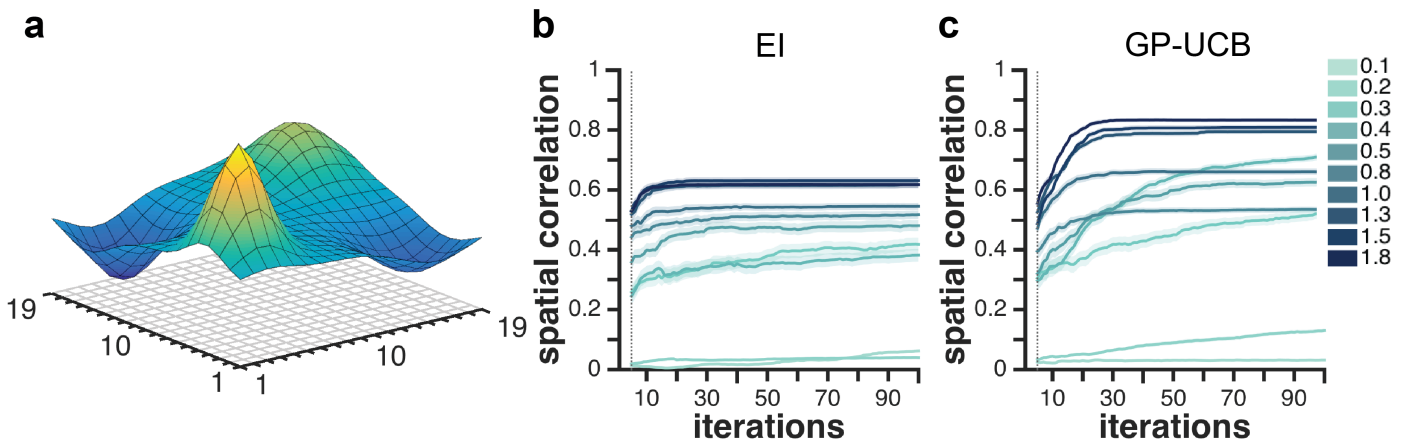


Fig. 3. Results of simulation analyses. (a) Modeled objective function used for simulations. Mean \pm SEM (shaded areas) spatial correlation between predicted and modeled parameter space for (b) EI and (c) GP-UCB acquisition function. As the first five iterations were used as a burn-in, they are not depicted here (gray dashed line). Simulations were performed for 10 different CNR values, ranging from 0.1 (bright blue) to 1.8 (dark blue).

algorithm to explore the parameter space over 100 iterations. The spatial correlation between the predictive posterior under the GP model and the true objective function was calculated at each iteration.

The mean \pm SEM spatial correlation results are depicted in Fig. 3b for the EI and in Fig. 3c for the GP-UCB acquisition function. Not surprisingly, we note that both acquisition functions fail in the presence of too much noise (for CNR values < 0.2). The results also indicate that the GP-UCB acquisition function clearly outperforms the EI acquisition function for typical CNR values reported in the literature (0.5-1.8).

IV. DISCUSSION AND FUTURE WORK

The results presented here serve as an initial exploration into the choices of kernel and acquisition functions which can be used in the context of real-time optimization for tACS. Our results propose the use of a squared exponential kernel function in combination with a GP-UCB acquisition function. In a series of simulations, the latter was shown to capture the whole parameter space (as identified by using spatial correlations) when compared to the EI acquisition function. This result was consistent for a range of CNR values tested. These findings serve as an important basis for upcoming real-time experiments. An avenue of future work is the definition of appropriate stopping criteria to determine convergence [15]. This is particularly important in our context due to high scanning costs, limited attentional capacities of subjects and neural habituation to the stimulation.

To our knowledge, this is the first study to propose a framework that allows tailoring non-invasive brain stimulation to an individual. The next step consists of applying this technology to patients suffering from traumatic brain injury; a condition that disrupts frontoparietal brain networks and hence results in cognitive impairments [16]. While closed-loop deep brain stimulation has already been shown to outperform the conventional approach in Parkinson's patients [17] we envision that our framework will advance personalized treatment by means of non-invasive brain stimulation. This will be of particular importance for neurological and psychiatric deficits that are diffuse and widely heterogeneous in their origin.

ACKNOWLEDGMENT

This research was supported by the NIHR Imperial BRC and the Wellcome Trust (103045/Z/13/Z).

REFERENCES

- [1] MF Kuo, MA Nitsche: Effects of transcranial electrical stimulation on cognition, *Clin. EEG Neurosci.*,43(3), 92–199 (2012)
- [2] CS Herrmann, S Rach, T Neuling, D Strueber: Transcranial alternating current stimulation: a review of the underlying mechanisms and modulation of cognitive processes, *Frontiers Human Neuroscience*, 7:279 (2013)
- [3] A Antal, W Paulus: Transcranial alternating current stimulation (tACS), *Frontiers Human Neuroscience*, 7:317 (2013)
- [4] R Lorenz, RP Monti, IR Violante et al.: The Automatic Neuroscientist: A framework for optimizing experimental design with closed-loop real-time fMRI. *NeuroImage*,129, 320–334 (2016)
- [5] E Brochu, VM Cora, N de Freitas: A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv:1012.2599* (2010)
- [6] B Shahriari, K Swersky, Z Wang, RP Adams, N de Freitas: Taking the Human Out of the Loop: A Review of Bayesian Optimization, *Proceedings of the IEEE*, 104(1), 148–175 (2016)
- [7] A Dove, S Pollmann, T Schubert, CJ Wiggins, DY von Cramon: Prefrontal cortex activation in task switching: an event-related fMRI study. *Cognitive Brain Research*, 9(1): 103–109 (2000)
- [8] AB Leber, NB Turk-Browne, MM Chun: Neural predictors of moment-to-moment fluctuations in cognitive flexibility, *PNAS*, 105(36), 13592–13597 (2008)
- [9] Y Koush, M Zvyagintsev, M Dyck, KA Mathiak, K Mathiak: Signal quality and Bayesian signal processing in neurofeedback based on real-time fMRI, *NeuroImage*, 59(1) (2012)
- [10] KJ Friston: Functional and Effective Connectivity: A Review, *Brain Connectivity*, 1(1),13–36 (2011)
- [11] CE Rasmussen, CK Williams: Gaussian processes for machine learning. MIT Press (2006)
- [12] D Duvenaud, JR Lloyd, R Grosse, JB Tenenbaum, Z Ghahramani.: Structure discovery in nonparametric regression through compositional kernel search, *ICML* (2013)
- [13] JT Lloyd, D Duvenaud, R Grosse, JB Tenenbaum, Z Ghahramani.: Automatic construction and natural-language description of nonparametric regression models, *Association for the Advancement of Artificial Intelligence* (2014)
- [14] M Welvaert, Y Rosseel: On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data, *PLoS One*, 8, doi: 10.1371/journal.pone.0077089 (2013)
- [15] R Lorenz, RP Monti, IR Violante et al.: Stopping criteria for boosting automatic experimental design using real-time fMRI with Bayesian optimization, *arXiv:1511.07827* (2015)
- [16] SR Jilka, G Scott, T Ham, A Pickering, V Bonnelle, RM Braga, R Leech, DJ Sharp: Damage to the Saliency Network and Interactions with the Default Mode Network, *J. Neurosci.*, 34(33), 0798-10807 (2014)
- [17] S Little, A Poghosyan, S Neal et al.: Adaptive deep brain stimulation in advanced Parkinson disease, *Ann. Neurol.*, 74(3), 449–457 (2013)