Robust Bayesian Optimization for Reliable FES Calibration

Keisuke Onoue
NAIST
Nara, Japan
onoue.keisuke.ok2@naist.ac.jp

Satoshi Endo
TU Munich
Munich, Germany
s.endo@tum.de

Kazushi Ikeda
NAIST
Nara, Japan
kazushi@is.naist.jp

Sandra Hirche
TU Munich
Munich, Germany
hirche@tum.de

***Abstract*—Functional Electrical Stimulation (FES) has been recognized as a potential approach for restoring fine motor control of the hand after central nervous system lesions, yet achieving precise selective control remains a challenge due to subject-specific anatomical variability. Conventional automated calibration procedures for high-density electrode arrays typically rely on exhaustive searches over numerous electrode combinations, rendering the setup process physically demanding and prone to inducing early muscle fatigue. Furthermore, standard modeling approaches fail to account for the heteroskedastic nature of FES-induced responses, where physiological noise varies significantly between active motor points and inexcitable regions. In this paper, we propose a robust calibration framework based on Gaussian Process regression with a Student-t likelihood to explicitly model heteroskedastic noise. We employ a Bayesian Optimization scheme using the Maximum Variance Reduction acquisition function to perform pure exploration within a strictly limited evaluation budget. To enhance spatial correlation learning, electrode pairs are modeled as an undirected graph and transformed into continuous geometric features. Experimental evaluations based on exhaustive search datasets from healthy participants reveal that standard Gaussian assumptions lead to severe predictive instability during the exploration of active motor regions. In contrast, the proposed Student-t model demonstrates robust convergence and superior log-likelihood scores. The results indicate that subject-specific response maps can be reliably characterized within 25 to 50 trials, or equivalent to three to four minutes, enabling the practical clinical deployment of complex multi-electrode FES systems within routine rehabilitation workflows.**

***Index Terms*—Functional Electrical Stimulation, Gaussian Process, Bayesian Optimization, Student-t Likelihood**

I. Introduction

Upper limb motor impairments, particularly those resulting in the loss of fine finger control, are prevalent consequences of central nervous system lesions such as stroke. Functional Electrical Stimulation (FES) has been widely recognized as a potential approach for restoring these lost motor functions by artificially triggering muscle contractions via low-level electrical pulses [1], [2]. While FES systems have demonstrated the potential to facilitate functions such as grasping in certain patient populations, achieving precise selective control over individual fingers remains a significant challenge due to the complex and overlapping musculoskeletal anatomy of the forearm [3]. Furthermore, as neuromuscular dynamics and anatomical structures are highly subject-specific, the FES model must be accurately calibrated for each individual to ensure effective and safe intervention.

Historically, FES interventions have required experts to manually identify optimal stimulation sites. This trial-and-error approach is not only time-consuming but frequently fails to achieve sufficient selectivity, leading to inefficient contractions and early muscle fatigue. To overcome these issues, recent research has transitioned towards high-density electrode arrays that allow dynamic switching of stimulation sites [1]. Automated calibration procedures have been proposed to identify optimal stimulation parameters using motion or force feedback, yet many existing systems rely on sequential search algorithms across all electrode pairs [1], which can still be time-consuming. Consequently, there is an urgent need for intelligent search strategies capable of characterizing subject-specific response maps within a strictly limited evaluation budget.

Biomechanical simulations and musculoskeletal models provide a theoretical foundation for predicting kinematic responses to FES [4], [5]. These models can simulate the effects of electrical fields on target muscles to identify optimal stimulation regions. However, a significant "Reality Gap" persists between these high-fidelity simulations and physiological reality [6], characterized by substantial aleatoric uncertainty arising from subject-specific factors, such as anatomical and morphological variations and electrode-skin contact dynamics. From a statistical perspective, these factors introduce complex non-linearities and heteroskedasticity [7]. As noted by Enoka et al. [8], FES recruitment order is relatively random and frequently contaminated by unintended sensory feedback loops. This results in "heavy-tailed" noise distributions where observations near sensitive motor points exhibit high-frequency fluctuations driven by spinal reflex activation, differing significantly from the stable background regions. Consequently, first-principles models alone are insufficient for reliable calibration [6]. Reliable calibration therefore requires robust, data-driven methods that explicitly model this non-uniform and heavy-tailed noise to distinguish between true physiological signals and measurement outliers [9], [10].

In this paper, we propose a robust calibration framework based on Gaussian Process (GP) regression [11] with a Student-t likelihood [12], [13] to explicitly model the heteroskedastic and heavy-tailed noise inherent in neuromus-

cular responses. This choice of likelihood ensures that the resulting response map is globally accurate and sufficiently robust against unmodeled stochasticity. The proposed model is integrated within a Bayesian Optimization (BO) [14], [15] scheme to guide the search process. While many BO applications employ reward-seeking strategies such as Expected Improvement [16] to find a single optimum, relying on a solitary operating point is risky in clinical practice. Robust clinical intervention requires a comprehensive understanding of the response surface to identify broad regions of effective stimulation that remain stable despite anatomical shifts or muscle fatigue. Consequently, we aim to reliably characterize regions of maximally selective motor activation, providing the essential foundation for effective, subject-specific FES calibration protocols. To achieve this, we employ a pure exploration strategy using the Maximum Variance Reduction (MVR) acquisition function. MVR sequentially selects inputs that maximize posterior uncertainty, which has been shown to be effective for minimizing simple regret R_N within a fixed evaluation budget N [17]. Here, assuming a maximization problem of an objective function f, simple regret is defined as: $R_N = f(x^*) - \max_{t=1,\dots,N} f(x_t)$, where $x^* = \arg\max_{x\in\mathcal{X}} f(x)$ is the global maximizer.

The efficacy of this MVR exploration strategy, however, hinges on accurate uncertainty estimation, which is fundamentally compromised in conventional GP-based FES modeling. Standard GP models typically assume a homoskedastic observation model where noise variance is constant across the entire input space [18], [19]. In contrast, FES data is inherently heteroskedastic; measurements in physiologically "silent" regions exhibit near-zero variance, whereas regions near motor points show high variance due to factors such as spinal reflex activation and muscle fatigue. To address this, we adopt a hierarchical approach that treats the Student-t distribution as a scale mixture of Gaussians. By assigning a latent auxiliary variable with a conjugate Gamma prior to each data point, the model estimates an individual observation precision for every trial. This structure enables the model to adaptively "down-weight" noisy measurements during inference, as the expected precision is naturally reduced for points with large residuals or high variability. To maintain the computational speed required for clinical practice, we employ Coordinate Ascent Variational Inference (CAVI) [20] to approximate the non-conjugate posterior. This allows for a computationally efficient characterization of non-uniform noise, ensuring that the predictive variance used by the MVR function remains a reliable indicator of model uncertainty rather than being inflated by transient physiological artifacts.

We validate the proposed framework through retrospective simulations using dense ground-truth datasets collected from healthy participants. Our results demonstrate that while standard Gaussian models exhibit significant instability when encountering high-variance motor responses, the Student-t likelihood effectively attenuates the influence of these physiological artifacts. This robustness enables the reliable identification of optimal stimulation sites within 25 to 50 trials. Consequently, the proposed approach allows for stable calibration in approximately three to four minutes, significantly enhancing clinical feasibility by preserving time and muscle capacity for subsequent rehabilitation.

II. Preliminaries

A. Gaussian Process Regression

Consider a regression problem with a dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$. We assume the observation model $y_i = f(\mathbf{x}_i) + \varepsilon_i$, where $f(\mathbf{x})$ is a latent function and ε_i represents the measurement noise. To infer the underlying function from noisy observations, we place a GP prior on $f(\mathbf{x})$:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')),$$

where $m(\mathbf{x})$ is the mean function, often assumed to be zero, and $k(\mathbf{x}, \mathbf{x}')$ is the covariance function (kernel) encoding the smoothness and structure of the function [11]. Following Bayes' theorem, under the noise assumption $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$, the predictive mean and covariance for a test input $\mathbf{x}_*$ is given as:

$$\begin{aligned} \mu_n(\mathbf{x}_*) &= \mathbf{k}_*^\top (K + \sigma_\varepsilon^2 I)^{-1} \mathbf{y}, \\ \sigma_n^2(\mathbf{x}_*) &= k_{**} - \mathbf{k}_*^\top (K + \sigma_\varepsilon^2 I)^{-1} \mathbf{k}_*, \end{aligned} \tag{1}$$

where K is the covariance matrix with $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{k}_* = [k(\mathbf{x}_1, \mathbf{x}_*), \dots, k(\mathbf{x}_n, \mathbf{x}_*)]^\top$, and $k_{**} = k(\mathbf{x}_*, \mathbf{x}_*)$.

B. Bayesian Optimization

BO is a sequential model-based approach designed for the global optimization of expensive-to-evaluate black-box functions f [14], [15]. The procedure consists of two primary components: a Bayesian surrogate model, typically a GP, and an acquisition function. At each iteration n, the surrogate model is updated with all available data to provide a posterior distribution, and the acquisition function guides the search by selecting the next evaluation point.

In this work, we prioritize global learning to facilitate safe optimization of the stimulation parameters. To this end, we employ the MVR acquisition function, a purely exploratory strategy that sequentially selects an input where the posterior uncertainty of the GP model is maximized [17]:

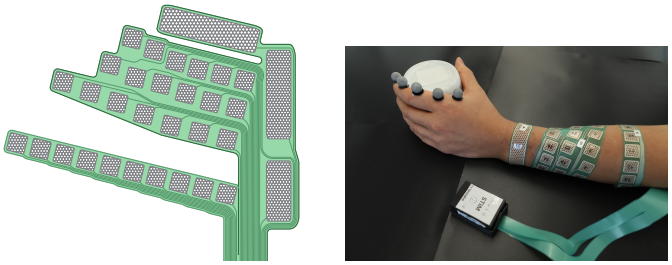
$$x_n = \arg\max_{x\in\mathcal{X}} \sigma_{n-1}^2(x). \tag{2}$$

By repeatedly sampling points with the largest predictive variance, the algorithm aims to reduce global uncertainty over the search space. After exhausting a fixed evaluation budget N, the framework returns a candidate maximizer based on the posterior mean of the learned GP model. However, in this work, we use the entire learned GP posterior as the calibrated subject-specific response map.

III. Methodology

A. Problem Statement

The primary objective of this framework is to efficiently learn the unknown mapping $y = f(\mathbf{x}) + \varepsilon$ between a stimulation pair $\mathbf{x} \in \mathcal{X}$ and the resulting motor response $y \in \mathbb{R}$ (e.g., flexion angle of the index finger). Given the



(a) Electrode layout. (b) Experimental setup.

Fig. 1: Electrode array and experimental setup: (a) Design and layout of the 32-channel high-density array; (b) Attachment of the array to the participant's forearm.

practical constraints of clinical settings, where the number of trials per patient (evaluation budget N) is strictly limited, the system must strategically explore the input space. While the ultimate clinical goal is to minimize simple regret to accurately identify the most effective stimulation sites for subsequent rehabilitation, the acquisition strategy prioritizes active learning to reduce global posterior uncertainty. This approach ensures that the resulting response map is globally accurate, providing a robust foundation for effective, subject-specific FES deployment.

B. Search Space Formulation as an Undirected Graph

We model the set of 32 electrodes as nodes $V = \{e_1, \ldots, e_{32}\}$ in a graph, arranged in the high-density grid structure illustrated in Fig. 1a. A single stimulation configuration $\mathbf{x}$ is defined as an undirected edge $\{e_i, e_j\}$ from this graph. The search space $\mathcal{X}$ corresponds to the edge set of a complete graph on $|V|$ nodes, resulting in $\binom{32}{2} = 496$ unique candidate stimulation pairs. This formulation assumes that, given the use of charge-balanced bi-phasic stimulation, swapping the anode and cathode does not significantly alter the functional muscle activation response.

C. Geometric Feature Engineering

To enable the GP model to capture spatial correlations, we transform each categorical electrode pair into a continuous 4-dimensional feature vector $\mathbf{z} \in \mathbb{R}^4$. Let $\mathbf{p}_i = (x_i, y_i)$ denote the 2D coordinates of electrode e_i on the flattened grid. For a pair $\{e_i, e_j\}$, we compute:

1) **Euclidean Distance:** $d = ||\mathbf{p}_j - \mathbf{p}_i||_2$, which influences the depth and volume of the induced electric field.
2) **Midpoint Coordinates:** $(m_x, m_y) = (\mathbf{p}_i + \mathbf{p}_j)/2$, representing the anatomical site of stimulation on the forearm in 2D.
3) **Normalized Angle:** $\theta = |\operatorname{atan2}(y_j - y_i, x_j - x_i)|$, normalized to $[0, \pi]$ to maintain uniqueness for undirected edges. This acts as a proxy for the orientation of the electric field relative to the underlying anatomy, allowing the kernel to learn anisotropic spatial correlations.

D. Response Map Creation via Pure Exploration

The motor response is modeled using GP regression described in Section II-A. To select subsequent stimulation candidates, we employ the MVR acquisition function described in Subsection II-B. This strategy separates exploration from exploitation, focusing first on minimizing model error across the entire search space. This global fidelity ensures that the subsequent identification of the optimal stimulation site is based on a reliable surrogate model rather than sparse or noisy data.

E. Modeling Latent Response under Heteroskedastic Noise

A significant challenge in FES calibration is that noise characteristics often vary across the input space, a phenomenon known as heteroskedasticity. For instance, stimulation in inexcitable regions typically yields zero or near-constant responses with minimal variance. In contrast, measurements in feasible motor regions exhibit higher variance due to physiological factors such as stochastic motor unit recruitment, spinal reflexes, and muscle fatigue. Accounting for this non-uniform noise is essential for maintaining predictive accuracy across the entire workspace. While standard approaches model noise variance using an auxiliary GP, these methods substantially increase the computational burden.

In this work, we address heteroskedasticity by adopting a Student-t likelihood, which offers inherent robustness against outliers and naturally accommodates data-dependent noise scales. The Student-t distribution can be represented as a scale mixture of Gaussians by introducing latent auxiliary variables $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_n]^\top$, which follow a Gamma distribution characterized by shape and rate parameters:

$$y_i \mid f_i, \lambda_i \sim \mathcal{N}(y_i; f_i, \sigma_\varepsilon^2/\lambda_i), \quad \lambda_i \sim \text{Gam}(\lambda_i; \nu/2, \nu/2),$$

where each λ_i serves as a local precision weight for the i-th observation, and $\text{Gam}(\nu/2, \nu/2)$ denotes the Gamma distribution with both shape and rate parameters set to $\nu/2$. The parameter ν dictates the degrees of freedom of the marginal Student-t distribution. As $\nu \to \infty$, the model approaches a Gaussian; lower values yield heavier tails. We selected comparatively small ν values to specifically ensure robustness against the heavy-tailed, sporadic noise characteristic of spinal reflexes.

Using a Student-t likelihood results in a non-conjugate model when paired with a GP prior, making the exact posterior analytically intractable. To address this while preserving computational efficiency, we employ CAVI. We use the factorized variational posterior

$$q(\mathbf{f}, \boldsymbol{\lambda}) = q(\mathbf{f}) \prod_{i=1}^{n} q(\lambda_i),$$

where $q(\mathbf{f}) = \mathcal{N}(\mathbf{f}; \mathbf{m}, \mathbf{S})$ and $q(\lambda_i) = \text{Gam}(\lambda_i; \alpha_i, \beta_i)$. The optimal local precision $\bar{\lambda}_i = \mathbb{E}_{q(\lambda_i)}[\lambda_i]$ is updated iteratively as:

$$\bar{\lambda}_i = \frac{\nu + 1}{\nu + \sigma_\varepsilon^{-2}((y_i - m_i)^2 + S_{ii})},$$

where m_i and S_{ii} denote the i-th element of variational mean vector $\mathbf{m}$ and the i-th diagonal element of the covariance matrix $\mathbf{S}$, representing the variational mean and variance of f_i, respectively. By fixing $\bar{\lambda}_i$ at each iteration, the model simplifies to a standard GP with heteroskedastic Gaussian noise $\mathbf{\Sigma}_\varepsilon = \text{diag}(\sigma_\varepsilon^2/\bar{\lambda}_1, \ldots, \sigma_\varepsilon^2/\bar{\lambda}_n)$. The kernel hyperparameters θ can then be optimized by maximizing the marginal log-likelihood (MLL):

$$\ln p(\mathbf{y} \mid \bar{\boldsymbol{\lambda}}, \theta) \propto -\frac{1}{2}\mathbf{y}^\top (K_\theta + \mathbf{\Sigma}_\varepsilon)^{-1}\mathbf{y} - \frac{1}{2}\ln |K_\theta + \mathbf{\Sigma}_\varepsilon|.$$

This alternating update scheme between variational parameters and kernel hyperparameters ensures fast and robust inference of the response map. The predictive distribution can also be obtained by replacing the constant noise in Equation 1 with the weighted noise.

F. Summary of the Proposed Method

We summarize the proposed calibration framework in Algorithm 1. This procedure integrates the graph-based search space formulation, geometric feature engineering, and robust GP modeling to identify optimal stimulation sites within a limited clinical budget. By using the MVR acquisition function, the framework prioritizes global uncertainty reduction to ensure that high-response areas are reliably characterized for deploying a complex FES system.

Algorithm 1 Proposed FES Calibration Framework

Require: Electrode set V, Budget N, Mapping $\Phi : \mathcal{X} \to \mathcal{Z}$
Search Space Construction:
  Define $\mathcal{X}$ of undirected edges from V
Feature Engineering:
  Compute features $\mathbf{z} = [d, m_x, m_y, \theta]$ for all $\mathbf{x} \in \mathcal{X}$
Initialization:
  Acquire initial dataset $\mathcal{D}_{n_0} = \{(\mathbf{z}_i, y_i)\}_{i=1}^{n_0}$
for $n = n_0 + 1, \ldots, N$ **do**
  Robust GP Inference:
    Update Student-t GP posterior approximation
  Acquisition Optimization (MVR):
    Select $\mathbf{z}_n = \arg\max_{\mathbf{z} \in \mathcal{Z}} \sigma_{n-1}^2(\mathbf{z})$
  FES Evaluation:
    Stimulate $\mathbf{x}_n$; observe response y_n
  Data Augmentation:
    $\mathcal{D}_n \leftarrow \mathcal{D}_{n-1} \cup \{(\mathbf{z}_n, y_n)\}$
end for
Calibration Result:
return Learned GP model

IV. Evaluation

To validate the proposed calibration framework, we conducted experiments using real-world FES data collected from two healthy participants, both of whom were male with an average age of 25. The evaluation aimed to (1) confirm the heteroskedastic nature of FES responses and (2) quantify the robustness and efficiency of the proposed GP with Student-t likelihood model compared to a standard Gaussian GP baseline.

A. Data Acquisition and Processing

1) Hardware Setup: We used a programmable FES stimulator (Tecnalia Research & Innovation) with a 32-channel electrode array. Stimulation consisted of biphasic pulses (25 Hz, 300 μs pulse width). The array was secured to the dorsal aspect of the forearm, overlying the extensor muscle group.

2) Response Metric: Kinematic data were recorded at 100 Hz using a Qualisys optical motion capture system with markers placed on the index finger joints, specifically the metacarpophalangeal (MCP), proximal interphalangeal (PIP), and distal interphalangeal (DIP) joints. The motor response Y was defined as the sum of the maximum angular displacements of these joints from their resting positions:

$$Y = \sum_{j \in \{\text{MCP, PIP, DIP}\}} (\theta_j(t_{peak}) - \theta_j(t_{start})).$$

3) Data Collection via Exhaustive Search: For each participant, we performed an exhaustive search of all $M = \binom{32}{2} = 496$ candidate electrode pairs. This search space assumes that the direction of the charge-balanced pulses does not significantly impact muscle recruitment, allowing us to treat electrode pairs as undirected edges. Each pair was evaluated 5 times to capture response variability, resulting in a total of 2,480 measurements per participant, conducted over a five-day period.

As illustrated in Fig. 1b, the participant's hand was fixed to a cylindrical object while maintaining a relaxed posture. Each individual stimulation sequence followed a standardized 5 s timing sequence:

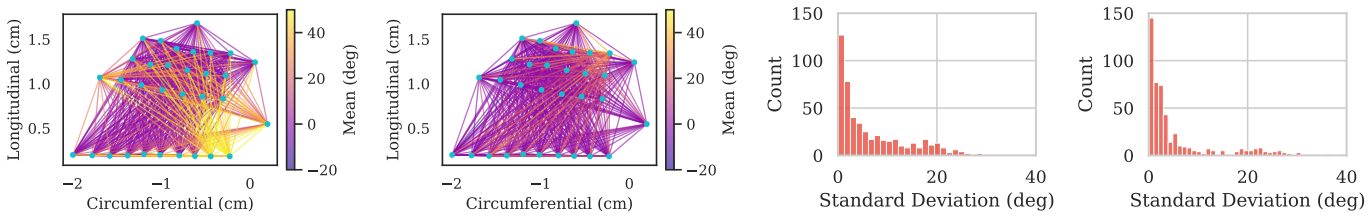
1) **Initial rest**: A baseline period of 1 s.
2) **Ramp-up**: A gradual increase of stimulation amplitude over 0.5 s.
3) **Stimulation**: Maintenance of maximum amplitude for 2 s.
4) **Ramp-down**: A gradual decrease of amplitude over 0.5 s.
5) **Final rest**: A recovery period of 1 s.

B. Evaluation Protocol

The efficacy of the proposed calibration framework was evaluated through simulated BO runs using the collected ground-truth dataset. For each participant, we conducted 50 independent runs to ensure that the reported results are representative of the framework's typical behavior rather than being artifacts of a specific random seed. Each run was assigned a total evaluation budget of $N = 100$ evaluations, consisting of:

1) **Initialization**: 1 sample selected via random sampling to initialize the surrogate model.
2) **BO-guided Search**: 99 iterations sequentially selected by the MVR acquisition function.

In the BO evaluation, we employed a resampling strategy to simulate realistic heteroskedastic noise. Instead of using a



(a) Subject 1 (b) Subject 2 (c) Subject 1 (d) Subject 2

Fig. 2: Ground-truth characteristics of the FES search space. (a)–(b) Spatial visualization of the mean response maps for Subject 1 and Subject 2, obtained by averaging measurements from five independent exhaustive searches. The response denotes the total angular displacement from the resting posture (0°); negative values indicate minor kinematic fluctuations observed during high-intensity contractions. (c)–(d) Histograms showing the distribution of empirical response standard deviations. The co-existence of low-variance (silent) and high-variance (active) regions confirms strong heteroskedasticity. Additionally, the distribution exhibits the effect of heavy-tailedness, supporting the use of the Student-t likelihood to robustly accommodate such variability.

TABLE I: Evolution of Predictive Log-Likelihood (PLL) on held-out test data. Values represent Mean ± Standard Deviation across 50 independent runs. Bold entries denote the best-performing model at each iteration. The proposed Student-t models (particularly $\nu = 3$) demonstrate superior stability and higher likelihood scores during the early exploration phase (Trials 5–25), whereas the Gaussian baseline exhibits extreme instability as indicated by the large standard deviations.

		Iteration					
Sub.	Model	5	10	15	25	50	100
S1	$\nu = 3$	$\mathbf{-3.107 \pm 1.843}$	$\mathbf{-1.965 \pm 0.621}$	$\mathbf{-1.904 \pm 0.707}$	$\mathbf{-1.728 \pm 0.412}$	-1.694 ± 0.370	-1.589 ± 0.221
	$\nu = 7$	-4.413 ± 3.283	-2.577 ± 1.268	-2.101 ± 0.644	-1.935 ± 0.559	$\mathbf{-1.560 \pm 0.178}$	-1.541 ± 0.141
	$\nu = 11$	-5.889 ± 4.770	-2.823 ± 1.533	-2.512 ± 0.940	-2.117 ± 0.785	-1.646 ± 0.167	-1.660 ± 0.117
	$\nu = 15$	-6.193 ± 5.607	-2.793 ± 1.267	-2.718 ± 0.923	-2.366 ± 0.777	-1.705 ± 0.252	-1.783 ± 0.116
	$\nu = 19$	-7.674 ± 7.417	-3.678 ± 2.805	-3.191 ± 1.586	-2.607 ± 0.877	-1.776 ± 0.254	-1.945 ± 0.142
	Gaussian	-321.193 ± 561.502	-168.528 ± 594.984	-24.272 ± 43.646	-8.637 ± 33.807	-1.582 ± 0.203	$\mathbf{-1.358 \pm 0.087}$
S2	$\nu = 3$	$\mathbf{-4.348 \pm 2.135}$	$\mathbf{-3.043 \pm 1.528}$	$\mathbf{-2.453 \pm 1.311}$	$\mathbf{-1.819 \pm 1.245}$	$\mathbf{-1.005 \pm 0.696}$	$\mathbf{-0.825 \pm 0.562}$
	$\nu = 7$	-6.986 ± 4.159	-4.922 ± 2.706	-4.105 ± 1.958	-3.156 ± 1.617	-2.280 ± 1.309	-1.820 ± 0.569
	$\nu = 11$	-9.379 ± 5.944	-6.313 ± 4.320	-5.026 ± 2.477	-4.125 ± 2.036	-3.334 ± 1.300	-2.726 ± 0.617
	$\nu = 15$	-11.642 ± 7.480	-7.545 ± 4.955	-5.927 ± 3.066	-5.118 ± 2.284	-3.390 ± 0.917	-3.055 ± 0.735
	$\nu = 19$	-13.559 ± 9.177	-8.887 ± 6.088	-7.066 ± 3.764	-6.324 ± 2.905	-4.226 ± 1.368	-3.559 ± 0.786
	Gaussian	-552.663 ± 2176.276	-155.871 ± 698.534	-9.838 ± 11.295	-8.791 ± 13.013	-4.434 ± 3.310	-2.040 ± 0.810

fixed mean value, the observation y_n for a selected pair $\mathbf{x}_n$ was randomly sampled from the five repeated samples available in the ground-truth. This strategy preserves natural physiological variability, capturing the aleatoric uncertainty driven by stochastic recruitment and reflex loops. Consequently, silent pairs exhibit near-zero variance, while active motor points display the high-variance, heavy-tailed distributions characteristic of clinical FES data, providing a realistic testbed for evaluating the robust Student-t likelihood.

To assess the global predictive capability, we monitor the mean Predictive Log-Likelihood (PLL) on held-out test data. At iteration t, the PLL is computed across the entire search space of $M = 496$ pairs as:

$$\mathrm{PLL}_t = \frac{1}{M} \sum_{j=1}^{M} \frac{1}{|\mathcal{H}_j^{(t)}|} \sum_{y \in \mathcal{H}_j^{(t)}} \log p_t(y \mid \mathbf{x}_j).$$

Here, p_t is the predictive distribution, and the test set $\mathcal{H}_j^{(t)}$ strictly excludes the training samples used in $\mathcal{D}_t$:

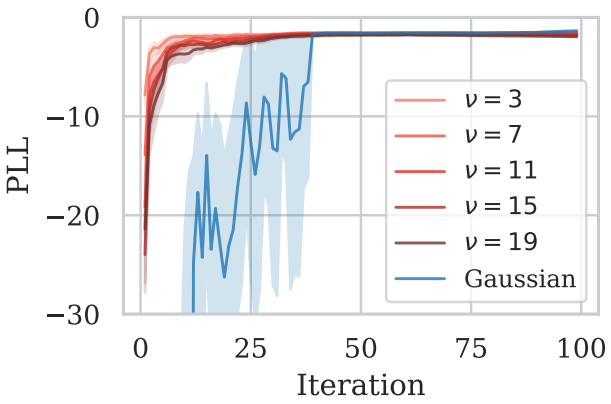
$$\mathcal{H}_j^{(t)} = \begin{cases} \{y_j^{(k)}\}_{k \neq k^*}, & \mathbf{x}_j \in \mathcal{D}_t \\ \{y_j^{(k)}\}_{k=1}^{5}, & \mathbf{x}_j \notin \mathcal{D}_t \end{cases},$$

where k^* denotes the index of the specific measurement sampled during the BO process.

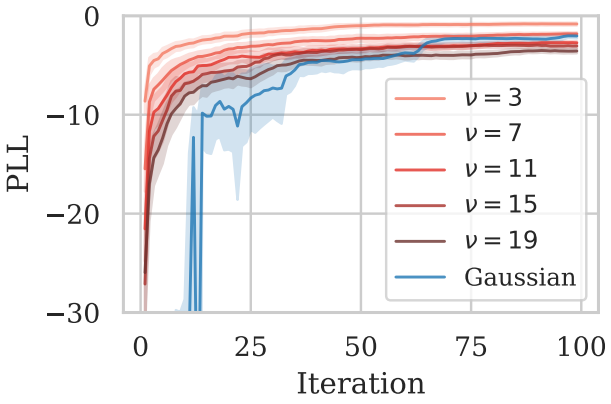
C. Results

1) Characteristics of FES Response: Fig. 2 presents the ground-truth landscape. Fig. 2(a)–(b) show the spatial localization of effective stimulation sites. Crucially, the variance histograms in Fig. 2(c)–(d) reveal a heavy-tailed distribution. The coexistence of low-variance (silent) regions and high-variance (active) regions confirms the strong heteroskedasticity of the data, validating the necessity of our hierarchical noise model.

2) Calibration Performance: Table I summarizes the results, and Fig. 3 illustrates the optimization trajectory. The proposed GP with Student-t models (particularly with $\nu = 3$) demonstrated superior stability compared to the Gaussian baseline. As shown in Fig. 3, the Gaussian GP (blue) exhibited extreme instability during the early exploration phase (Trials 1–40), characterized by sharp drops in likelihood. This indicates that the Gaussian model failed to handle high-variance outliers, degrading its global predictive capability. In contrast, the Student-t models maintained a monotonic improvement.



(a) Subject 1



(b) Subject 2

Fig. 3: Evolution of predictive performance during the calibration process. The plots show the Predictive Log Likelihood (PLL) on held-out test data for (a) Subject 1 and (b) Subject 2 over 100 trials. The proposed GP with Student-t likelihood models (red gradients, $\nu \in \{3, \ldots, 19\}$) demonstrate rapid and stable convergence from the onset. In contrast, the standard Gaussian GP baseline (blue) exhibits significant instability during the early exploration phase (e.g., Iterations 1–40 in Subject 1), characterized by sharp drops in likelihood when encountering high-variance observations. Shaded regions denote the standard error across 50 independent runs.

Quantitatively, for Subject 2 (Table I), the Student-t model ($\nu = 3$) consistently outperformed the Gaussian baseline, achieving a final PLL of -0.825 versus -2.040.

V. Discussion

The results demonstrate the critical importance of explicitly modeling heteroskedastic noise to ensure stable FES calibration. As the MVR acquisition function actively seeks uncertainty, it naturally drives the search toward high-variance regions. However, in a physiological context, such variance is frequently exacerbated by factors such as unintended spinal reflexes rather than informative signal trends A standard Gaussian likelihood interprets this physiological variance as model error, inflating global noise and destabilizing the map. In contrast, the Student-t model mitigates this by down-weighting high-variance observations via the latent precision variable λ, enabling robust learning even with sparse data.

A. Clinical Utility and Best Spot Identification

Clinically, identifying the stable "best spot" for stimulation is paramount. Although MVR targets global uncertainty reduction rather than direct maximization, the rapid convergence of the posterior mean (Fig. 3) confirms that the framework can reliably pinpoint optimal sites. Furthermore, the Student-t model's robustness against outliers reduces "false positives," preventing the mistaken selection of noisy regions that appear as peaks due to unmodeled stochasticity.

B. Methodological Considerations

Several factors remain to be addressed for further refinement. Regarding the geometric representation, our flattened 2D features simplify the actual 3D forearm structure. This projection creates a risk where the distance between electrodes at the wrapping boundaries is overestimated, potentially degrading the spatial correlation learning of the GP model. To address this geometric distortion, incorporating geodesic distances on a cylindrical manifold would better model the 3D anatomy and mitigate these edge effects. Similarly, we treated electrode pairs as undirected edges based on the assumption of biphasic symmetry. However, if polarity shifts significantly affect recruitment due to nerve orientation, this simplification could mask distinct directional responses. Future studies should verify these effects to ensure that the search space definition fully captures the physiological reality. Furthermore, we assumed a time-invariant mapping throughout the procedure. While fatigue was likely minimal during our short calibration window, this assumption limits applicability in longer sessions where physiological states drift. Long-term applications must account for such response shifts caused by muscle fatigue to maintain the accuracy of the optimal parameters. Finally, the current implementation is limited to a discrete pairwise search. Extending this approach to simultaneous multi-electrode activation, often referred to as virtual electrodes, could offer a more continuous stimulation space. This expansion would allow the system to identify optimal stimulation points that lie between physical electrodes, thereby potentially improving spatial resolution.

C. Physiological Modeling and Selectivity

Our method treats neuromuscular physiology as a "black-box" mapping. This data-driven approach is robust to the "Reality Gap" inherent in biomechanical simulations because it avoids reliance on rigid anatomical assumptions. However, ignoring prior knowledge means the model must learn the response surface entirely from data, which places a heavier burden on the exploration phase. Adopting a "gray-box" approach could accelerate calibration by incorporating known physiological features, such as recruitment curves or reflex dynamics, directly into the model structure. Moreover, clinical

neurorehabilitation demands precise selectivity across fingers rather than simple contraction. While the current method maximizes single-digit flexion, functional tasks require coordinated control to avoid unintended grasping synergies. The framework can be extended to multi-objective optimization to address this need. This extension would allow the system to maximize target activation while simultaneously minimizing the unintended co-activation of adjacent digits.

D. Clinical Feasibility and Time Efficiency

The clinical adoption of advanced FES systems is currently hindered by the time-intensive nature of the calibration procedure. Compared to conventional methods ($\approx 5\,\mathrm{min}$), our framework adds negligible inference overhead ($\approx 5\,\mathrm{s}$/iteration). As the Student-t model achieves stable accuracy between Iterations 25 and 50 (Table I), calibration can be completed in $3\,\mathrm{m}$ to $4\,\mathrm{m}$. This efficiency supports early stopping, offering a significant advantage for clinical deployment.

VI. Conclusion and Future Work

We presented a robust FES calibration framework that employs Gaussian Process regression with a Student-t likelihood to explicitly model heteroskedastic neuromuscular noise. Integrated with a Maximum Variance Reduction acquisition strategy, our approach demonstrates superior predictive stability compared to standard Gaussian baselines, effectively handling physiological outliers. Experimental results confirm that subject-specific response maps can be reliably characterized in approximately three to four minutes. This significant reduction in setup time minimizes patient fatigue, offering a clinically viable solution for a use of complex FES system in personalized neurorehabilitation. Future work will extend this framework to multi-objective optimization to further enhance fine finger selectivity.

Ethics

Ethical approval for the study was obtained from the Ethics Committee of the Technical University of Munich (approval no. 355/20 SSR). All participants provided written informed consent prior to participation.

Acknowledgment

This work was supported by the European Research Council (ERC) Consolidator Grant "Safe data-driven control for human-centric systems (CO-MAN)" under grant agreement No. 864686, and by the Innovation Network eXprt of the Technical University of Munich (TUM), funded by the Federal Ministry of Research, Technology and Space (BMFTR) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the States. This work was partially supported by the Robotics Institute of Germany (RIG), funded by the BMFTR.

References

[1] H. Usman, Y. Zhou, B. Metcalfe, and D. Zhang, "A functional electrical stimulation system of high-density electrodes with auto-calibration for optimal selectivity," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8833–8843, 2020.

[2] C. Höhler, S. Endo, J. Hermsdörfer, L. Cazenave, H. Kavianirad, K. Jahn, and C. Krewer, "Repetitive grasping after stroke assisted by functional electrical stimulation," *Artificial Organs*, 2025.

[3] H. Kavianirad, F. Missiroli, S. Endo, L. Masia, and S. Hirche, "Toward dexterous hand functional movement: Wearable hybrid soft exoglove-fes study," in *2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 2024, pp. 1346–1351.

[4] D. C. McFarland, B. I. Binder-Markey, J. A. Nichols, S. J. Wohlman, M. De Bruin, and W. M. Murray, "A musculoskeletal model of the hand and wrist capable of simulating functional tasks," *IEEE transactions on biomedical engineering*, vol. 70, no. 5, pp. 1424–1435, 2022.

[5] J. Baier, S. Selkmann, and B. Bender, "Simulation of fes on the forearm with muscle-specific activation resolution," *Frontiers in Bioengineering and Biotechnology*, vol. 12, p. 1384617, 2024.

[6] A. Sena, H. Kavianirad, S. Endo, E. Burdet, and S. Hirche, "The gap in functional electrical stimulation simulation," in *3rd Workshop on Closing the Reality Gap in Sim2Real Transfer for Robotics*, 2022.

[7] J. S. Petrofsky, H. J. Suh, S. Gunda, M. Prowse, and J. Batt, "Interrelationships between body fat and skin blood flow and the current required for electrical stimulation of human muscle," *Medical engineering & physics*, vol. 30, no. 7, pp. 931–936, 2008.

[8] R. M. Enoka, I. G. Amiridis, and J. Duchateau, "Electrical stimulation of muscle: electrophysiology and rehabilitation," *Physiology*, 2019.

[9] J. R. Medina, H. Börner, S. Endo, and S. Hirche, "Impedance-based gaussian processes for modeling human motor behavior in physical and non-physical interaction," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 9, pp. 2499–2511, 2019.

[10] N. Das, S. Endo, H. Kavianirad, and S. Hirche, "Framework for learning a hand intent recognition model from semg for fes-based control," in *2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 2024, pp. 1320–1327.

[11] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[12] J. Vanhatalo, P. Jylänki, and A. Vehtari, "Gaussian process regression with student-t likelihood," *Advances in neural information processing systems*, vol. 22, 2009.

[13] P. Jylänki, J. Vanhatalo, and A. Vehtari, "Robust gaussian process regression with a student-t likelihood." *Journal of Machine Learning Research*, vol. 12, no. 11, 2011.

[14] P. I. Frazier, "A tutorial on bayesian optimization," *arXiv preprint arXiv:1807.02811*, 2018.

[15] R. Garnett, *Bayesian optimization*. Cambridge University Press, 2023.

[16] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," *Advances in neural information processing systems*, vol. 24, 2011.

[17] S. Vakili, N. Bouziani, S. Jalali, A. Bernacchia, and D.-s. Shiu, "Optimal order simple regret for gaussian process bandits," *Advances in Neural Information Processing Systems*, vol. 34, pp. 21 202–21 215, 2021.

[18] E. M. Schearer, Y.-W. Liao, E. J. Perreault, M. C. Tresch, W. D. Memberg, R. F. Kirsch, and K. M. Lynch, "Semiparametric identification of human arm dynamics for flexible control of a functional electrical stimulation neuroprosthesis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 12, pp. 1405–1415, 2016.

[19] E. M. Schearer and D. N. Wolf, "Predicting functional force production capabilities of upper extremity functional electrical stimulation neuroprostheses: a proof of concept study," *Journal of Neural Engineering*, vol. 17, no. 1, p. 016051, 2020.

[20] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.