

Introducing simulation-based inference to c-VEP BCI: A feasibility study

Sébastien Velut^{1,2}, Marie-Constance Corsi³, Sylvain Chevallier², Frédéric Dehais¹, Jordy Thielen⁴

¹Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, Toulouse, 31055, France

²INRIA TAU - LISN - Université Paris-Saclay, Gif-sur-Yvettes, 91190, France

³CNRS, Inria, Inserm, AP-HP, Hopital de la Pitié Salpetriere, Paris, 75013, France

⁴Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, 6525 GD, Netherlands

E-mail: sebastien.velut@isae-supero.fr

ABSTRACT: Code-modulated visual evoked-potential (c-VEP) based reactive brain-computer interfaces (BCIs) deliver high information-transfer rates with minimal calibration, yet performance collapses when generalizing across participants. Additionally, only few studies focus on understanding the neural mechanisms underlying c-VEP, which remain poorly understood. In this paper, we introduce the simulation based inference (SBI) framework to the BCI field. SBI allows to offload calibration on simulated synthetic data generated with parameters that may represent a large cohort of different participants. We assess SBI's feasibility to provide competitive classification performance on empirical data as well as inferring response parameters, while being calibrated only on synthetic data. We achieved 82% classification accuracy with a within participant setup using a c-VEP based BCI, and a mean squared error of 0.11 between the empirical visual evoked potential and the inferred one. These results suggest the feasibility of SBI for BCI, and could deepen our understanding of the mechanisms underlying the c-VEP. We present the current limitations of the SBI framework and give future directions to address these.

INTRODUCTION

Among reactive brain-computer interfacing (BCI) paradigms, a recent paradigm, known as the code-modulated visual evoked potential (c-VEP), involves using pseudo-random noise sequences for visual stimulation. Such c-VEP based BCIs have gained more and more attention due to their high information-transfer rates, robustness, and reduced calibration requirements [1].

Despite these advantages, several challenges remain. First, performances often decrease when generalizing between sessions or participants due to inter/intra-subject variability in the EEG. This variability induces the need for calibration in every usage of a BCI. Second, the generative mechanisms underlying the c-VEP, that is the forward model, are still poorly understood, and studies typically rely on discriminative machine-learning models, that is backward models, such as template matching [1] or Riemannian methods associated with linear discriminant analysis [2]. In turn, explainability and physiologi-

cal modeling remain largely underexplored.

To address these challenges, a forward model, like recon-volution [3] can be used to perform inference, recovering response parameters given observed data. Such inference is not trivial, as there can be an infinite number of combinations of parameters that may have caused the observation, that is the problem is ill-defined.

A candidate solution is to use Bayesian methods such as Markov chain Monte Carlo (MCMC) [4] or variational inference [5]. Despite its robustness in arriving at good solutions, such algorithms typically rely on explicit likelihood evaluations, which are computationally expensive or even infeasible in a real-time system such as a BCI [6]. As an alternative to such iterative optimization strategies, simulation-based inference (SBI) [7, 8] embraces a forward model to generate a quasi infinite amount of labeled training data over a wide range of potential parameters. Then, an inverse model can be calibrated to recover these parameters based on the generated synthetic data. If successful, such an SBI framework yields a powerful inverse model that can recover parameters from any arbitrary empirical data, after an initial training period on synthetic data. This so-called amortization could relieve a BCI from its tedious calibration session.

SBI is a field within machine learning that is getting more and more attention because of its wide range of applications, such as in astrophysics and particle physics, and more recently in neuroscience [9–11].

SBI has not yet made its entry into the BCI field, despite its potential to mitigate or even solve several of the large challenges in the BCI field. Particularly, amortized inference could be useful in addressing the inter/intra-subject variability. Additionally, SBI may provide interpretable parameters improving the explainability of the EEG signal and understanding of underlying neural mechanisms. However, SBI depends on several factors that could impact its performance such as the simulation fidelity, the nature of the inference, the nature of the embedding net used before the inference, and the prior specification used to create simulated data [8], which raises non-trivial challenges for EEG applications.

In this study, we investigate the feasibility of applying SBI to c-VEP-based BCI decoding. Specifically, we

showcase the potential of an SBI framework and identify key challenges along the way. Doing so, we analyze both *simulated-to-simulated* and *simulated-to-empirical* scenarios to assess the adequacy and generalization behavior of the model. Beyond performance evaluation, this work explores the potential of SBI as a bridge between decoding accuracy and neural interpretability in c-VEP BCIs. This study does not aim to outperform existing c-VEP classifiers. Instead, we evaluate whether SBI constitutes a viable modeling framework for c-VEP responses by jointly assessing classification performance, parameter recovery, and model adequacy.

MATERIALS AND METHODS

Dataset: We used a previously collected empirical dataset called Ricker patch-based StAR c-VEP (ISAE-SUPAERO Dataset [12]), which received approval from the ethics committee of the University of Toulouse (CER approval number 2023-749) and was conducted in compliance with the Declaration of Helsinki. The EEG data were collected using a dry 8-electrode Enobio system, with a sampling rate of 500 Hz, positioned over the occipital and parieto-occipital sites (PO7, O1, Oz, O2, PO8, PO3, POz, PO4).

This dataset comprises 24 healthy individuals (4 women, 20 men, mean age = 29.3 years, SD = 7.5 years). Participants were instructed to direct their gaze to one of five disks indicated by a 0.5 s cue, after which the disks flashed for 2.2 s according to unique pseudo-random noise-codes. Each participant performed 15 runs, which consisted of 5 trials, one for each of the disks in random order. In summary, the dataset contains a total of 75 trials of 2.2 s, including 15 trials of per class ($N = 5$).

Pre-processing: First, we apply a finite impulse response bandpass filter between 1 Hz and 45 Hz. As the stimulus-onset asynchrony of individual flashes within the stimulus sequences exceeds 0.35 s, we create overlapping epochs of size 0.35 s every 1/500 s (the EEG sampling rate). This gives $E = 15 * (2.2 - 0.35) * 500 = 13875$ epochs per participant per class. Second, to ensure comparable amplitude over participants, we normalize the data by dividing the data by the standard deviation of each participant's training set. Third, to improve the signal-to-noise ratio (SNR) of the captured event-related potential (ERP) response to flashes of the c-VEP, and improve the quality of the signal, we add a spatial filtering step using xDAWN [13]. We select the 4 highest-ranked xDAWN components [14].

To align the different participants into the same feature space and reduce the variability between the subjects, we apply a Riemannian spatial alignment, an effective method for cross-subject analysis [15]. However, to perform the alignment, we first need to create super epochs by combining the average of the spatially filtered epoch with each spatially filtered target response. Let $\mathbf{E}_Z = \frac{1}{E} \sum_{k=1}^E \mathbf{Z}_k$ be the evoked response of the spatially filtered data where E is the number of epochs and

$\mathbf{Z}_k$ is the k^{th} epoch of the spatially filtered target data. $\mathbf{Z}_k \in \mathbb{R}^{C \times T}$, where C is the number of electrode channels, and T is the number of time samples in one epoch. We create the super epochs where the k^{th} sample will be $\mathbf{Z}'_k = \{Z_{k,1}, Z_{k,2}, Z_{k,3}, Z_{k,4}, E_{Z,1}, E_{Z,2}, E_{Z,3}, E_{Z,4}\}$. Combining this preserved information with the application of the spatial filter, intra-epoch variability is mitigated. We compute the evoked response of the spatially filtered data using the 22 s calibration data, and we concatenate it to all epochs, including both the training and test sets.

Then, to perform the Riemannian spatial alignment, we use the participant's calibration phase as a training batch and compute the associated covariance matrices with the Ledoit-Wolf estimator [16]. Then, we measure the Riemannian mean of these matrices, and finally, we transport the samples of the participant with this mean as follows:

$$\hat{\mathbf{Z}}'_i = \bar{\mathbf{R}}^{-\frac{1}{2}} \mathbf{Z}'_i \quad \forall i \in [1, \dots, E] \quad (1)$$

where $\mathbf{Z}'_i \in \mathbb{R}^{C \times T}$ is an element of $\{\mathbf{Z}'_i\}_{i \in [1, \dots, E]}$, $\bar{\mathbf{R}}$ is the Riemannian mean of the covariances matrices of the set $\{\mathbf{Z}'_1, \dots, \mathbf{Z}'_A\}$ and A is the number of samples in the calibration set of the participant. As the forward model presented in the next sections output data with only one channel, we select, at the end, only the first components of the preprocessed data.

Simulation-based inference: For SBI framework we need to define three specific components (see Figure 1). Firstly, we define a forward model that can generate c-VEP responses given a stimulus sequence and a set of VEP parameters. Secondly, we define a prior over the VEP parameters to sample from and input to the forward model and generate a synthetic dataset. Finally, we define a backward model that can retrieve the VEP parameters from any c-VEP data. This backward model composed of an embedding net and a density estimator to retrieve the parameters from the data.

To perform the training of the SBI framework, we first create simulated trials, 15000 trials in our study, by sampling the parameters from the prior and using the forward model with a specific stimulus sequence to generate data. This yields a synthetic dataset of c-VEP responses given ground-truth VEP parameters. Subsequently, we train the backward model (embedding net and density estimator) on this synthetic dataset and its ground-truth labels.

Once trained, we evaluate the performance of the SBI pipeline to regress the parameters given an unseen synthetic dataset, and to classify the attended code sequence given a hold-out empirical dataset.

The SBI framework was implemented using the `sbi` Python library [17]. The following sections describe the three components of the SBI framework in more detail.

SBI Forward Model: The forward model generates an EEG response to a given c-VEP stimulus making use of the deconvolution framework [3]. Specifically, we assume that the c-VEP $s_i(t)$ can be modeled as a convolution of the i -th stimulus $c_i(t)$ and a temporal response function $r(t)$:

$$s_i(t) = c_i(t) * r(t) \quad (2)$$

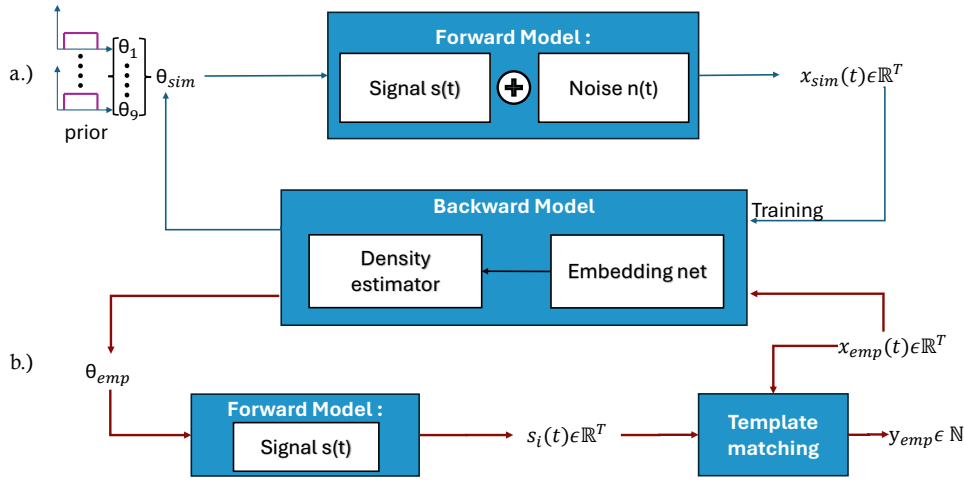


Figure 1: **Schematic of the SBI framework.** (a.) The training loop that samples parameters, generates data using the forward model, uses that to train a backwards model to regress these parameters. (b.) The classification procedure on empirical data to retrieve the attended code sequence. N : Number of codes, T : number of timesteps, θ : parameters, $x(t)$: EEG data, $s_i(t)$: templates, y : labels, $_{sim}$: simulated, $_{emp}$: empirical.

The stimulus sequence $c_i(t)$ is one of the 5 pseudo-random binary noise-codes used during the experiment. We model the temporal response function $r(t)$, which represents the flash-VEP, as the addition of 3 Gaussian components, which ideally represent the N75, P100, and N125 complex of a classical flash-VEP. Each Gaussian component is modeled with 3 parameters: the amplitude, mean and standard deviation of the Gaussian, corresponding to the amplitude, latency and width of the peak. Finally, we add measurement noise $n(t) \sim \mathcal{N}(0, \sigma)$ represented by a zero-mean Gaussian with variance $\sigma = 0.1$ to obtain the simulated data $x_{sim}(t)$:

$$x_{sim}(t) = s_i(t) + n(t) \quad (3)$$

Thus, the forward model is parameterized by a set of 9 parameters θ_{sim} characterizing the underlying VEP model $r(t)$. Given a set of 9 parameters θ_{sim} and a stimulus code $c_i(t)$, the forward model outputs simulated c-VEP source activity $x_{sim}(t)$ of size T , with T corresponding to the number of time-points in our data.

SBI Backward Model: The backward model takes EEG data and retrieves the VEP parameters. It is composed of two components, an embedding net and a density estimator. The aim of the embedding net is to reduce the dimensionality of the input and to create features for the density estimator. The embedding net is important as the density estimator needs clear, informative and low-dimensional features. The density estimator yields a posterior distribution over the VEP parameters from the features created by the embedding net.

As embedding net we used a recurrent neural network (RNN) composed of one hidden layer with 128 units followed by a linear layer with 20 units and a ReLU activation layer. This was motivated by a preliminary

benchmark performed between four different embedding nets: a handcrafted summary statistic function, a convolutional neural network, the encoding part of a trained auto-encoder, and the RNN. The RNN was selected as it performed best on the simulated data when performing the local classifier two-sample test (LC2ST) [18].

As density estimator we used the sequential neural posterior estimator (SNPE) [19]. SNPE is the most widely used density estimator for SBI and can learn the likelihood function from compressed representations of the data. SNPE trains an inference network to directly approximate the posterior distribution $p(\theta|x)$, using samples (θ, x) drawn from the joint distribution $p(\theta, x)$. Unlike methods that focus on predicting a single point estimate of the parameters, SNPE trains the network to predict the parameters of a probability distribution over θ , conditioned on the observation sample. Such networks are trained by minimizing the conditional negative log-likelihood of simulated parameters under the neural posterior estimator. Once the inference network is trained, it can be used to infer θ from any new observed data $x_{obs}(t)$, providing the approximate posterior distribution $p(\theta, x_{obs}(t))$.

A key advantage of this approach is that the network, once trained, can perform inference efficiently for any new observations, either synthetic or empirical, without needing additional simulations or retraining. This amortization feature makes SBI powerful for analyzing many different observations from the same model. This gives a special advantage compared to time-consuming algorithms like MCMC sampling.

SBI Prior Parameters: The SBI framework requires a prior distribution over the VEP parameters to generate synthetic data used to train the backward model. Ideally,

this prior distribution represents the expected range of parameters for the unknown target distribution, here the empirical data.

We define the prior for each parameter as a uniform distribution, limited to a certain range. We informed the lower and higher boundary of this range by analysing the empirical data. That is, for each participant, we fit the VEP model $r(t)$ of the forward model on averaged target epochs. From this fit, we retrieve the 9 parameters, and proceed in one of the following two ways. First, to model a prior across participants, we set the range to the smallest and largest value found across participants, referred to as the *across-pp prior*. Second, to model a prior per individual participant, we apply a small range around the best individual parameters (± 0.02 for the latency, ± 0.2 for the amplitude, ± 0.001 for the width), referred to as the *within-pp prior*.

Analysis: The SBI framework is trained by sampling from the prior, generating synthetic data, and calibrating the backward model using that synthetic data. After training, the SBI framework is evaluated in two ways.

First, we evaluate the regression of the VEP parameters using an unseen simulated dataset. For this analysis, we restrict to synthetic data, because we have access to the ground truth parameters, which is not the case for the empirical dataset. For this analysis, we report the root mean squared error (RMSE).

Second, we evaluate the SBI framework by performing a classification analysis. That is, we apply the backwards model to retrieve VEP parameters and use these to generate template responses using the signal component of the forward model $s(t)$. Specifically, we apply the forward model given each of the stimulus sequences c_i , with $i \in [1, \dots, 5]$, and the predicted parameters θ , but refrain from adding any noise $n(t)$ to the signal $s(t)$ to generate templates $s_i(t)$. Then, we can perform a standard template-matching classifier that maximizes the correlation between data $x(t)$ and templates $s_i(t)$ to retrieve the code index i :

$$\hat{y} = \arg \max_i \text{corr}(x(t), s_i(t)) \quad (4)$$

For each participants, the classification analysis was performed for both prior on both an unseen synthetic dataset of 100 trials, as well as the empirical dataset composed of 75 trials, referred to as *simulated-to-empirical* and *simulated-to-simulated*, respectively.

RESULTS

To assess the performance of retrieving the VEP parameters, we first compute the root mean squared error between the ground-truth parameters used to create the synthetic data in the *simulated-to-simulated* mode for both priors and the associated retrieved parameters. The results are described in Table 1. We see that within-pp RMSE is lower than across-pp RMSE.

Secondly, we assess the performance of retrieving the VEP parameters by computing the mean squared error

between the ground truth VEP response and the inferred VEP response. For the *simulated-to-simulated*, we obtain an average value of 0.08 ± 0.07 for the within-pp prior and 0.29 ± 0.26 for the across-pp prior. For the *simulated-to-empirical*, we get an average value of 0.11 ± 0.06 for the within-pp prior and 0.45 ± 0.17 for the across-pp prior.

Figure 2a shows the VEPs from the *simulated-to-empirical* analysis. We observed that the within-pp prior fits the empirical VEP very well, even for participant P21 which showed the highest MSE value. However, we observed that the VEP responses created with the parameters retrieved from the SBI pipeline trained with the across-pp prior, are far from the empirical value and can explain the low performance with the across-pp prior.

We further assessed the SBI performance by estimating classification accuracy. Figure 3 shows the classification accuracy of both priors in both synthetic to synthetic, as well as synthetic to empirical modes. We observed that the within-pp prior reached a significantly higher accuracy than the across-pp prior (86.2% vs 33.8%, $p < .001$). Moreover, the *simulated-to-simulated* performance is significantly higher than the *simulated-to-empirical* (36.8% vs 30.8% for the across-pp prior, $p = .037$; 89.8% vs 82.6% for the within-pp prior, $p = .028$).

DISCUSSION

To the best of the authors' knowledge, this study is the first to apply simulation-based inference (SBI) in the field of BCI, specifically to the c-VEP paradigm. SBI offers the potential to eliminate the initial calibration step that is typically required for classification models, since, within this framework, training can be performed on a virtually unlimited amount of simulated data.

In this feasibility study, we showed that satisfactory performance can be achieved both in recovering the underlying response parameters and in application-level decoding accuracy, provided that the prior parameters are chosen appropriately. Specifically, priors defined at the individual-participant level yielded high classification accuracy, whereas a universal across-participant prior led to a substantial decrease in performance.

The lower accuracy observed with the across-participant prior may be explained by several neurophysiological factors, including inter-subject and intra-subject variability [20]. In particular, the VEPs estimated from the empirical data showed that not all participants exhibited the canonical three-peak flash-VEP pattern, consisting of a negative peak followed by a positive peak and then a second negative peak [2, 12]. In addition, the latencies of these peaks varied considerably across several participants. This variability may have made it difficult to define a broad across-participant parameter range and also led to overlap in the latency ranges of the three components.

While we aimed to evaluate the full SBI framework, at two stages we relied on empirical data to estimate properties of the framework itself. This warrants caution in

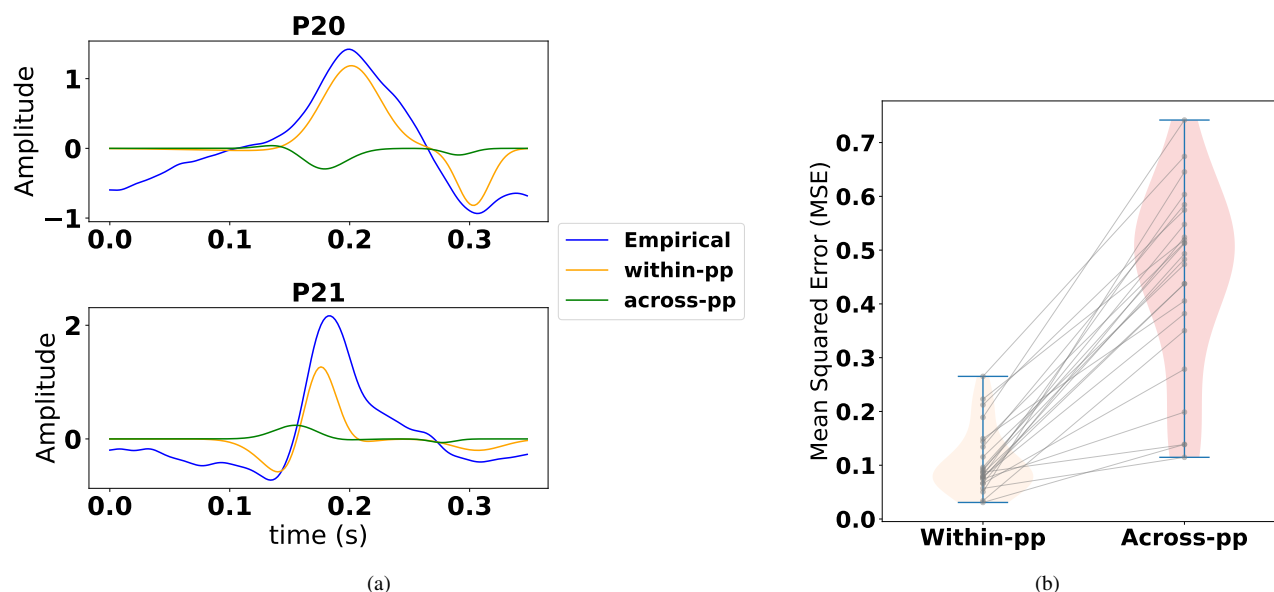


Figure 2: **Regression performance of the SBI framework.** (a) Displayed are the empirical visual evoked potential (VEP) (blue line), predicted VEP using the within-participant prior (yellow line), and predicted VEP with the across-participant prior (green line). Two participants are shown: on the top the participant with the lowest MSE and on the bottom the participant with the highest MSE. (b) Shown is the MSE between the empirical VEP and the simulated VEP created from the parameters retrieved with the within-pp prior (peach color) or with the across-pp prior (coral color). Each dot corresponds to the MSE of one participant. Line connect the points of the same participant.

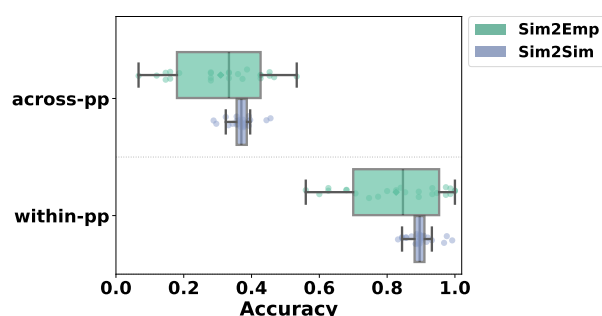


Figure 3: **Classification performance of the SBI framework.** Accuracy for the *simulated-to-simulated* (Sim2Sim) and *simulated-to-empirical* (Sim2Emp) modes for the across-participant and within-participant priors.

the interpretation of the results and limits the extent to which the current framework can be generalized.

First, we simulated a single EEG component extracted with xDAWN and applied Riemannian alignment. This improved performance compared with modeling a single channel. While it required calibration of xDAWN and the Riemannian alignment using only 22 s of data, it violates the amortization of SBI, being dependent on some empirical data for the calibration process.

Second, also the prior distributions were inspired by the empirical data, which under the true SBI framework would not be fully known. Especially constraining to a participant-specific prior considerably improved results. Future work should improve the across-participant setting, as well as test the generalization beyond the currently used dataset, EEG setup, and stimulus protocol. To avoid contradictory information in the training of the

SBI framework, we used only one code when creating the simulated dataset to train the SBI. However, when using a broader prior like the across-participant one, the classification steps tend to classify the data as the same label as the one used in for the training of the SBI, explaining the low performance for the across-pp prior.

In a preliminary analysis, we used the across-pp prior and trained the SBI using all 5 codes, which increased the *simulated-to-simulated* accuracy to 81.3%, but the accuracy remained at 32% for the *simulated-to-empirical*. Future work should integrate all stimulation codes into the SBI training procedure to improve generalization across codes in empirical data. In addition, it will be important to investigate whether the SBI framework can be extended to other experimental protocols.

The forward model used in this study can be considerably improved to reflect realistic EEG data. First, the current model enforces a 3-peak flash-VEP, while some participants showed fewer or more peaks, allowing further finetuning. Second, the current model is fully linear, not modeling any non-linear dynamics like caused by habituation. Third, the noise model considered measurement noise only, ignoring known EEG properties like $1/f$ noise, potential artifacts such as eye blinks, etc. While the current results already show feasibility, it is expected that an improved forward model may further improve results. In the backward model, we used an RNN for the embedding net. We believe that this embedding net could be improved by optimizing its hyperparameters or by developing a new embedding net that better captures features for the density estimator in the c-VEP domain. Furthermore, future work should aim at making the backward

Table 1: **Root mean squared error.** RMSE between predicted parameters and ground-truth used to create synthetic data in the *simulated-to-simulated* mode for both priors.

	Amplitude (a.u.)			Latency (s)			Width (a.u.)		
	N75	P100	N125	N75	P100	N125	N75	P100	N125
across-pp	3.9e-6 ±	9.5e-6 ±	5.6e-6 ±	9.3e-2 ±	7.2e-1 ±	1.3e-1 ±	1.8e-7 ±	3.5e-8 ±	4.1e-8 ±
	3.6e-6	9.6e-6	3.6e-6	8.8e-2	4.1e-1	9.1e-2	1.4e-7	2.6e-8	3.4e-8
within-pp	1.3e-7 ±	1.4e-7 ±	2.2e-7 ±	1.3e-3 ±	1.5e-3 ±	1.6e-3 ±	1.0e-8 ±	7.0e-9 ±	8.1e-9 ±
	8.9e-8	1.1e-7	9.7e-8	9.3e-4	9.8e-4	1.1e-3	4.6e-9	6.0e-9	6.8e-9

model invariant to the stimulus protocol.

While the across-participant setting showed poor performance, the within-participant setting showed promising results. An interesting future direction could be to start the SBI with an uninformed prior for any new user. Subsequently, the prior can be updated by using the predicted parameters from empirical data of the current user. The updated prior can then initiate the calibration of a new SBI, more tailored to that current user. That is, we can perform an active learning SBI framework that adapts in real-time, ultimately improving its performance on unseen data.

CONCLUSION

This study is the first of its kind to demonstrate the feasibility of applying the simulation based inference (SBI) framework in c-VEP based BCI. We show good accuracy in classification of empirical data while having trained the framework on synthetic data. We observed that a prior across participants decreased overall accuracy as compared to a prior within participants, in line with substantial inter-subject variability. We discussed several challenges and how to improve the SBI framework in future work, paving the way for retrieving response parameters and classification of empirical data, while offloading calibration to synthetic data.

REFERENCES

[1] Martinez-Cagigal V, Thielen J, Santamaria-Vazquez E, Perez-Velasco S, Desain P, Hornero R. Brain-computer interfaces based on code-modulated visual evoked potentials (c-VEP): A literature review. *Journal of Neural Engineering*. 2021.

[2] Cabrera Castillos K, Ladouce S, Darmet L, Dehais F. Burst c-VEP based BCI: Optimizing stimulus design for enhanced classification with minimal calibration data and improved user experience. *NeuroImage*. 2023;11.

[3] Thielen J, Broek P van den, Farquhar J, Desain P. Broad-Band visually evoked potentials: Re(con)volution in brain-computer interfacing. *PLOS One*. 2015.

[4] Gilks WR, Richardson S, Spiegelhalter D, Eds. *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC: New York (1995).

[5] Zhang C, Butepage J, Kjellstrom H, Mandt S. *Advances in Variational Inference*. Oct. 23, 2018.

[6] Csilléry K, Blum MGB, Gaggiotti OE, François O. Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*. 2010;410–418.

[7] Deistler M et al. “Simulation-Based Inference: A Practical Guide.” arXiv. Tech. Rep. 2025.

[8] Lueckmann JM, Boelts J, Greenberg DS, Gonçalves PJ, Macke JH. Benchmarking Simulation-Based Inference. 2021.

[9] Brehmer J. Simulation-based inference in particle physics. *Nature Reviews Physics*. 2021.

[10] Dax M, Green SR, Gair J, Macke JH, Buonanno A, Schölkopf B. Real-Time Gravitational Wave Science with Neural Posterior Estimation. *Physical Review Letters*. 2021.

[11] Bernardo D et al. Simulation-based inference of developmental EEG maturation with the spectral graph model. *Communications Physics*. 2024.

[12] Dehais F, Cabrera Castillos K, Ladouce S, Clisson P. Leveraging textured flickers: A leap toward practical, visually comfortable, and high-performance dry EEG code-VEP BCI. *Journal of Neural Engineering*. 2024.

[13] Rivet B, Souiloumiac A, Attina V, Gibert G. xDAWN algorithm to enhance evoked potentials: Application to brain-computer interface. *IEEE Transactions on Biomedical Engineering*. 2009.

[14] Aristimunha B et al. *Mother of all BCI Benchmarks*. Version 1.4.3. 2025.

[15] Mellot A, Collas A, Chevallier S, Gramfort A, Engemann DA. Geodesic optimization for predictive shift adaptation on EEG data. In: *NeurIPS*. 2024.

[16] Ledoit O, Wolf M. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*. 2004.

[17] Boelts J et al. sbi reloaded: A toolkit for simulation-based inference workflows. *Journal of Open Source Software*. 2025.

[18] Linhart J, Gramfort A, Rodrigues PLC. L-C2ST: Local Diagnostics for Posterior Approximations in Simulation-Based Inference. In: *NeurIPS*. 2023.

[19] Greenberg DS, Nonnenmacher M, Macke JH. Automatic Posterior Transformation for Likelihood-Free Inference. In: *ICML*. 2019.

[20] Velut S, Thielen J, Chevallier S, Corsi MC, Dehais F. Neurophysiological screening of individual variability for robust decoding in c-VEP-based BCI. *Imaging Neuroscience*. 2026.