Comments to the Author

1. Is the manuscript technically sound, and do the data support the conclusions?

The manuscript must describe a technically sound piece of scientific research with data that supports the conclusions. Experiments must have been conducted rigorously, with appropriate controls, replication, and sample sizes. The conclusions must be drawn appropriately based on the data presented.

Reviewer #1: Yes
Reviewer #2: Yes
Reviewer #3: Yes

---

2. Has the statistical analysis been performed appropriately and rigorously?

Reviewer #1: Yes
Reviewer #2: Yes
Reviewer #3: Yes

---

3. Does the manuscript adhere to the PLOS Data Policy?

Authors must follow the PLOS Data policy, which requires authors to make all data underlying the findings described in their manuscript fully available without restriction. Please refer to the author's Data Availability Statement in the manuscript. All data and related metadata must be deposited in an appropriate public repository, unless already provided as part of the submitted article or supporting information. If there are restrictions on the ability of authors to publicly share data—e.g. privacy or use of data from a third party— these reasons must be specified.

Reviewer #1: Yes
Reviewer #2: Yes
Reviewer #3: Yes

---

4. Is the manuscript presented in an intelligible fashion and written in standard English?

PLOS ONE does not copyedit accepted manuscripts, so the language in submitted articles must be clear, correct, and unambiguous. Any typographical or grammatical errors should be corrected at revision, so please note any specific errors here.

Reviewer #1: Yes
Reviewer #2: Yes
Reviewer #3: Yes

---

5. Review Comments to the Author

Please use the space provided to explain your answers to the questions above. You may also include additional comments for the author, including concerns about dual publication, research ethics, or publication ethics. (Please upload your review as an attachment if it exceeds 20,000 characters)

Reviewer #1:

Iturrate et.al "Unsupervised online self-calibration of error-related potentials for brain-machine interfaces".

This is strictly a methods paper, but does answer a fairly important scientific question. The methods and their presentation are adequate. The bibliography could include a few more references on the technical aspects of the methods being presented.

The authors offer an innovative contribution to self-calibrating brain-machine interface literature that involves a "look-ahead planning" approach to defining uncertainty. The problem is established quite elegantly in the first few paragraphs of the introduction. As for the proposed approach, it seems pretty straightforward. Very short summary: during the initial stages of training, there is uncertainty in terms of tasks and signals. This is minimized by including uncertainty into the model (a Bayesian minimization model). You might read through the description of the method a couple of time to ensure clarity, as it is quite complicated.

Code is located on Github (the readme instructions are quite useful), which is adequate for purposes of open-access and replicability. Can the data be made public as well? This might help readers along with the tutorial materials you presented (Github can be used to archive data matrices as well). While including all of the data might be superfluous, data for a few key intervals and/or subjects would be helpful.

Potential problem: could you get more subjects? Are eight subjects enough for a proof-of-concept? I say this because there appear to be individual differences in the data (see Table 1 and Figure 4). These might be captured by grouping the existing data by one or two performance parameters, and then used to interpret the outcomes of this larger experiment.

The focus of this approach addresses minimizes the time needed to calibrate BCIs. The presentation of this is clear enough. The formatting is adequate. However, a few points on the text and approach in general:

* for the first instance of "train ErrP classifier", define ErrP using the full name and cite a reference to the method.

* are the learning curves (percent accuracy) shown in Figure 6 different due to individual differences or features of the algorithmic approach? If it is due to individual differences, then are these features of the signal or behavioral performance? See my point above about acquiring more subject data.

Reviewer #2:

The paper describes a self-calibration procedure for BCIs used in grid-based reaching tasks. The paper is well written and I have only some minor suggested changes to improve the readability of the paper.

1. Since the procedure relies on the constraints from a grid reaching task, the title would be more informative if it mentioned that. As is, it seems a bit misleading that it can be applied to any brain-machine interface paradigm.
2. Authors mention data is fully available, but I only saw a link to the open source codes on github. This should be made clearer, as to adhere to PLOS One's open data policy.
3. The authors mention in line 39 that preliminary results have been presented in [19,20]. In order to clarify the novelty of the present paper relative to these previous works, a couple of sentences could be added to that effect.
4. Line 77-78, it is mentioned that data is bandpass filtered from 1-10Hz, then they mention downsampling to 32Hz. I am assuming the 10Hz is a typo and should in fact read 100Hz; correct?
5. One claim in line 138 is that the calibration procedure seems to be sensitive to the coherence of the EEG patterns for the same sets of labels. This was not given enough attention in the paper, and could strengthen the "practicality" aspect of the proposed calibration. Can the authors quantify the effect that breaking this assumption has on the results?
6. Figure 4b would be enhanced if the subject numbers were overlaid with the points. I would be interested in seeing whose points correspond to the curves on figure 6. Subjects 4 and 7, for example, have very similar curves - where are they in figure 4b?
7. Authors argue that the two curves in figure 5 are the same in line 373. Maybe a better claim is "insignificantly different"? There seems to be a clear latency difference between the two curves of about 25ms. Any potential explanation?

Regarding the text:

1. line 7, remove "the" from the large subject specificity
2. line 11, out-of-the-box use of *the* BCI
3. line 12: medical applications*, such* as those
4. line 31: have also achieved control (i.e., remove 'a')
5. line 56: replace 'and as incorrect' by 'or as incorrect'
6. line 336: replace 'have demonstrated' with 'having demonstrated'
7. line 373: careful with 'same' in 'same accuracy and grand average signals' (see comment 7 in section above)

Reviewer #3:

The authors present a new method for self-calibration of BCIs for reaching tasks demonstrated on a single 2D task and based on their previous work on the matter. The adaptive algorithm takes advantage of the task constraints to simultaneously learn the calibration parameters and the task.

The authors claim that their method could be employed in any other reaching task as long as the number of reaching positions is finite. The mathematical formulation seems to support the claim but practical issues relating convergence and complexity of the task should be discussed.

Some of the assumptions also seem controversial to me and should be further discussed:

- p3, l94: equally probable label a prioris. Couldn't any further information about this a priori be gathered from the task? In fact, in the task analysed in the experiments, this does not seem to be the case.

- p4, l109: it is assumed that the user follows optimal policies to reach the task but this is not consistent with the affirmation of the previous item ("we do not have a priori knowledge of the user intended assessment of the action"). Besides, will this also be true for more complicated tasks?

Some other comments:

- p2, l59: how do blinks and eyes motion affect the system? Is this related with the differences in performance observed for different subjects?

- p3, l86: 'one' should be 'an'.

- p3, l100: the variable 'r' is not defined. Does it stand for 'reward'?

- p4, l117: rephrase the sentence. Do you mean 'candidate tasks' versus 'the target task'?

- p3, l88, p5. l152, p6. l196: k multiple times defined.

- References should be fully revised. There are many errors related to missing information such as volumes, pages, capital letters (for example, 'EEG' instead of 'eeg').

- fig. 4a: vertical axis does not reflect percentages as the label states.