Reply to Reviewer's Comments on "EEG Waveform Analysis of P300 ERP with applications to Brain Computer Interfaces"

We are grateful to the reviewer for pointing out relevant issues in our manuscript. In the following, we discuss how we dealt with each raised issue.

· ·

REVIEWER #2 TRANSCRIPT:

This manuscript presents findings that contrast the performance of a selection of BCI classification methods in a single subject database. All in all, I find several aspects of this paper quite interesting, particularly its attempt to assess the impact of latency and component amplitude variation on classification performance. However, overall, I fear the range/scope of the paper is too limited to make a significant contribution to the literature.

1. Abstract, while automated analysis of EEG data certainly has significant potential with regards to the future of clinical EEG research, it is too soon to unqualifiedly state that these methods as they are outshine the research done by the clinical EEG community. Some aspects of the introduction also veer towards hyperbole on occasion and may be better received if they were more conservatively stated.

2. Several notable papers in the domain of P300-BCI work should be referenced (e.g., Rakotomamonjy & Guigue, 2008) as well as other prior efforts to evaluate P300-BCI classification methods (e.g., Manyakov, Chumerin, Combaz & Hulle, 2011; Oweis, Hamdi, Ghazali & Lwissy, 2013).

4. Page 12, Lines 383-394. It should be noted that the dataset from which the template ERP subject was drawn from was an ALS patient dataset. It is currently unclear why Subject Number 8 and trial number 2 was selected out of all the subjects and trials in the BNCI-Horizon database.

5. Integrating the template ERP signal into a Null-EEG stream and assuming everything is there except the P300 ERP component is probably an untenable assumption. As indicated in the manuscript itself, for instance, the null-EEG stream is a passive task. While the non-P300 trials in an active task contain additional activity associated with processes including target stimuli anticipation, active task focus, cognitive fatigue, etc. This is not true of the Null-EEG stream. Furthermore, combining and subsequently classifying EEG data from two different subjects is questionable at best given the physiological differences between the two. Differences in recording condition, equipment and task stimuli should also be considered. Collectively, these factors make it difficult to assess the value of the classification results observed on this study.

6. The separate manipulation of latency, as well as amplitude noise, is interesting and if extended may in and of itself prove to be an interesting manuscript in and of itself (e.g. contrasting the impact of a range of different latency variations and amplitude noise on BCI classification ability), true EEG data almost invariantly involves not only a mix of both noise sources but other sources such as cognitive fatigue, systematic task stimuli adaptation etc. Thus all three experiments represent assessments of isolated aspects of noise which would be interesting if expanded on in a more expansive simulation set (see point 4 and summary) but currently are too limited in scope to draw strong conclusions regarding the expected inter-component performance in real-world data.

It seems as though the entire study is built on the classification of a single active P300 BCI-Speller template from an ALS participant into a single null-EEG stream of a healthy subject performing a passive-viewing version of a BCI-Speller task. The generalizability of these results are thus difficult, if not

impossible to assess. Given that assessments of classification methods for P300 BCI interfaces are fairly common (e.g. Manyakov et al., 2011; Oweis et al. 2013; Krusienski et al. 2010), it is unclear what the value of another contrast of classification methods is to the field. The assessment of the impact of sources of noise is interesting (see point 6), but would need to be expanded on considerably to be of practical value. Furthermore, while this manuscript frequently emphasizes the value of BCI to the clinical field. However, most clinicians are generally uninterested in classifying the presence/absence of a P300 response, arguably the most obvious cognitive ERP component extant. Instead, the focus in that area is instead on subtler dynamics in the EEG signal such as frequency slowing, triphasic waves, and generalized ictal activity. Indeed, the introduction of this manuscript mentions several EEG components of value which notably does not include the P300 in its list. If demonstrating clinical utility is of interest, it would seem more appropriate to target the classification of these subtler elements with a larger sample set and broader response ERP pool.

General Comments

This manuscript presents findings that contrast the performance of a selection of BCI classification methods in a single subject database. All in all, I find several aspects of this paper quite interesting, particularly its attempt to assess the impact of latency and component amplitude variation on classification performance. However, overall, I fear the range/scope of the paper is too limited to make a significant contribution to the literature.

....

> We would like to honestly thank the Reviewer for this absolutely excellent and detailed review. We have modified the experiments and several parts of the manuscript in order to make it more suitable for publication.

....

1. Abstract, while automated analysis of EEG data certainly has significant potential with regards to the future of clinical EEG research, it is too soon to unqualifiedly state that these methods as they are "outshine the research done by the clinical EEG community." Some aspects of the introduction also veer towards hyperbole on occasion and may be better received if they were more conservatively stated.

....

> We apologize for this mistake. We definitely selected an incorrect wording because the message that we aimed to convey was exactly the opposite. We have modified the abstract to emphasize the message and the idea that we failed to transmit in our original manuscript.

....

2. Several notable papers in the domain of P300-BCI work should be referenced (e.g., Rakotomamonjy & Guigue, 2008) as well as other prior efforts to evaluate P300-BCI classification methods (e.g., Manyakov, Chumerin, Combaz & Hulle, 2011; Oweis, Hamdi, Ghazali & Lwissy, 2013).

....

> These important references were added to the manuscript. Thank you for pointing them out.

....

4. Page 12, Lines 383-394. It should be noted that the dataset from which the template ERP subject was drawn from was an ALS patient dataset. It is currently unclear why Subject Number 8 and trial number 2 was selected out of all the subjects and trials in the BNCI-Horizon database.

....

The reason we selected that particular trial and subject was because our manuscript stressed the importance of the waveform, the shape of the ERP. The obtained P300 waveform of Subject Number 8, trial 2 was the most prototypical that we've found along the BNCI-Horizon dataset of ALS patients. The analysis was performed by visual observation. We have now added also 4 more for the Active part of the pseudo-real dataset.

We have now included this reason on the manuscript which was mistakenly omitted on our first version.

....

5. Integrating the template ERP signal into a Null-EEG stream and assuming everything is there except the P300 ERP component is probably an untenable assumption. As indicated in the manuscript itself, for instance, the null-EEG stream is a passive task. While the non-P300 trials in an active task contain additional activity associated with processes including target stimuli anticipation, active task focus, cognitive fatigue, etc. This is not true of the Null-EEG stream. Furthermore, combining and subsequently classifying EEG data from two different subjects is questionable at best given the physiological differences between the two. Differences in recording condition, equipment and task stimuli should also be considered. Collectively, these factors make it difficult to assess the value of the classification results observed on this study.

....

The recording conditions of both datasets are exactly the same. The EEG device, though not the same exact model, is from the same manufacturer and the sampling frequency is as close as possible. All the preprocessing stage was implemented according to (Riccio *et al.*, 2013).

Following Reviewer comments, the pseudo-real dataset is now divided in two. First, we've replicated the experiment conducted by Riccio *et al.* (2013) et al on 4 healthy subjects performing an active-modality P300 copy-spelling task while using the P300-Based BCI Speller. We also extended the original single null-EEG stream to 4 subjects while performing the passive-modality where the letter to copy-spell is not informed to the user and the subject is unaware of how the matrix works.

Regarding the active-modality, for each subject, the most prototypical P300 waveform template is selected and extracted. This template is superimposed on each preprocessed segment marked as hit. The baseline was obtained by picking randomly any no-hit segment for the same subject.

....

6. The separate manipulation of latency, as well as amplitude noise, is interesting and if extended may in and of itself prove to be an interesting manuscript in and of itself (e.g. contrasting the impact of a range of different latency variations and amplitude noise on BCI classification ability), true EEG data almost invariantly involves not only a mix of both noise sources but other sources such as cognitive fatigue, systematic task stimuli adaptation etc. Thus all three experiments represent assessments of isolated aspects of noise which would be interesting if expanded on in a more expansive simulation set (see point 4 and summary) but currently are too limited in scope to draw strong conclusions regarding the expected inter-component performance in real-world data.

....

We appreciate a lot this insight which we also found interest to look after and analyze more in detail. However, we would like to clarify that the purpose of these experiments were to assess the incidence of the most common source of noise in the generation of the P300, to the shape of the obtained waveform of the P300, "how the averaged signal looked-like".

....

It seems as though the entire study is built on the classification of a single active P300 BCI-Speller template from an ALS participant into a single null-EEG stream of a healthy subject performing a passive-viewing version of a BCI-Speller task. The generalizability of these results are thus difficult, if not impossible to assess. Given that assessments of classification methods for P300 BCI interfaces are fairly common (e.g. Manyakov et al., 2011; Oweis et al. 2013; Krusienski et al. 2010), it is unclear what the value of another contrast of classification methods is to the field. The assessment of the impact of sources of noise is interesting (see point 6), but would need to be expanded on considerably to be of practical value. Furthermore, while this manuscript frequently emphasizes the value of BCI to the clinical field. However, most clinicians are generally uninterested in classifying the presence/absence of a P300 response, arguably the most obvious cognitive ERP component extant. Instead, the focus in that area is instead on subtler dynamics in the EEG signal such as frequency slowing, triphasic waves, and generalized ictal activity. Indeed, the introduction of this manuscript mentions several EEG components of value which notably does not include the P300 in its list. If demonstrating clinical utility is of interest, it would seem more appropriate to target the classification of these subtler elements with a larger sample set and broader response ERP pool.

....

We would like to emphasize that the purpose of this work was to raise awareness about the automatic procedures that can analyze EEG signals by their waveforms, mimicking what traditionally EEG physicians performed on the clinical EEG. As far as we are aware of, no work has addressed this topic in this form and we humbly believe it could be a contribution. The Reviewer is right, but the objective of this manuscript was to establish the connection, and to focus on the goal of the BCI discipline to foster interaction between the clinical and BCI community (Chavarriaga *et al.*, 2017).

We have modified the manuscript to emphasize these aspects.

....

REFERENCES

Riccio, A.; Simione, L.; Schettini, F.; Pizzimenti, A.; Inghilleri, M.; Belardinelli, M.O.; Mattia, D.; Cincotti, F. Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis. *Frontiers in Human Neuroscience* **2013**, *7*, 732.

Chavarriaga, R.; Fried-Oken, M.; Kleih, S.; Lotte, F.; Scherer, R. Heading for new shores! Overcoming pitfalls in BCI design. *Brain-Computer Interfaces* **2017**, *4*, 60–73.