Reply to Reviewer's Comments on "Training a gaming agent on brainwaves"

We appreciate a lot all the comments and feedback provided by the AE and the Reviewers.
In the following, we discuss all the changes with each raised issue.

. . . . . .

AE TRANSCRIPT:

AE's Comments to Author:
Associate Editor Comments to the Author:
The paper is improved but would benefit from a further update to

- improve abstract
- provide further interpretation/discussion of findings
- fix minor errors/typos
- address the grand average recommendation

REVIEWER #1 TRANSCRIPT:

Comments to the Author The paper is improved. The contribution is clearer. Thank you for the reply to comments. I attach some further comments.

Training a Gaming Agent on Brainwaves The paper has been improved. IMO, there are still areas that need further attention to improve the paper further. Page 1 The abstract still reads like an introduction until the last sentence. I would begin with sentence: "Error-related potential (ErrP) are a particular type of ERP that can be elicited by a person who attends a recognizable error."

Expand the results of the paper as this is the contribution. "Results show that there is an effective transfer of information and that the agent learns successfully to solve the game efficiently. Both the underlined terms need to cite evidence from the paper.

Line 40 "This information is used to make a gaming agent <that> improves ...."

Line 11 col 2 "how biological agents learn from its <their> environment by exploring it and getting feedback rewards, either negative or positive."

Line 17 col2 "~~Nonneglected~~ is the influence of DeepBrain's AlphaGo project," // Re-word

Line 24 col 2 "The papers [8], [9], [10] have successfully demonstrated that a robot can be controlled by obtaining a reward signal from a person's brain activity, ~~which~~ who is observing the robot, <to> solve a task."

Line 58 Put in a reference number from ethics committee from which approval was sough (e.g. University committee)

Page 2 Line 56 "Data ~~is~~ <are> handled and processed with the OpenVibe Designer," // datum-data

Line 37 col 2 "This dataset has been published on the IEEE DataPort initiative [15]." // I would move this to the end of next paragraph (or possibly end of paper)

Page 4 Lin 24 This allows ~~to learn~~ the Q-Table <to learn> based on the subject's feedback from the movements the agent took, which are chosen pseudo-randomly, while executing the brainwave session.

Line 50 "The best overall performance is obtained using Logistic Regression." // also need to discuss the significance of the overall levels of accuracy in Discussion. Is a best performance of 0.672 useful/acceptable? How does it compare to other researchers (if a comparison is possible).

Line 42 col 2 "These results are also consistent with their classification ROC curves, shown ~~on~~ <in> Figures 6 obtained for both subjects, where the area under the curve are close to chance level." // Why didn't you show one good ROC curve and one bad one. This would provide the basis for comparison.

Page 5 Figure 5 “Y axis shows the averaged number of steps, while x axis show the number of experiences used to cumulative train the Q-Table.” // Why does Fig5 E only have 2 sets of data? Some discussion of the (smaller) change from 1 – 2 would be appropriate (e.g. I Discussion) Line 51 “Not performance gain is evidenced, the agents learn nothing which implies that the reward information is useless.” // useless is not a good choice of wording –provides no value

Line 55 “It can be seen that the overall performance of the agent improves as long as there are more experiences to be used to train it, regardless if they were generated from the brainwaves classification from different subjects.” // more explanation/discussion needed. Is this an average effect from positive learners?

This work aims to state whether ErrP signals could be used to train a gaming agent using reinforcement learning. The collected data show that ErrP signals can in fact be classified and used to train an agent effectively. // Can you link these findings to confirm/challenge other recent research in Err potentials

Page 6 Fig 8 – why use data from subject 6 (a non-learner)?

Line 42 “However, even though this implies that the agent misses frequently that an action taken is wrong, this is not hindering the overall performance and the agent is still learning” //

Line 20 col 2 Despite that, the rewards generated from different subjects can be used to train the same Q- Table to improve its performance, which may lead to strategies where the overall performance is improved based on the information from different human critics at the same time. //These are key findings, worth more discussion/interpretation and inclusion in abstract.

REVIEWER #2 TRANSCRIPT:

Comments to the Author The authors have satisfactorily addressed my concerns. I have a minor comment regarding Figure 3. It is unclear what the brown bar labelled GrandAvg represents since there is one for each participant. Please clarify. I suggest having another set of bars (a ninth) for the grand average (average over participants) for each of the algorithms used.

AE General Comments

AE’s Comments to Author:

Associate Editor Comments to the Author: The paper is improved but would benefit from a further update to

- improve abstract

  We have rewrote the abstract explaining in further details the obtained results.

- provide further interpretation/discussion of findings

  We included extra information and verified the contributions/findings. This information was included in the abstract as well as in the Conclusion (Section IV).

- fix minor errors/typos

  We fixed the carefully reported issues gently provided by Reviewers.

- address the grand average recommendation

  We addressed this issue. We think there could be a misunderstanding from our side about the suggestion/recommendation of the Reviewer. We added additional information in this response letter, and apologize for this.

Reviewer 1 General Comments

Comments to the Author The paper is improved. The contribution is clearer. Thank you for the reply to comments. I attach some further comments.

The paper has been improved. IMO, there are still areas that need further attention to improve the paper further. Page 1 The abstract still reads like an introduction until the last sentence. I would begin with

sentence: “Error-related potential (ErrP) are a particular type of ERP that can be elicited by a person who attends a recognizable error.”

We started the abstract as suggested, and rewrote it detailing the obtained results. Thank you very much for your suggestion.

Expand the results of the paper as this is the contribution. “Results show that there is an effective transfer of information and that the agent learns successfully to solve the game efficiently. Both the underlined terms need to cite evidence from the paper.

We included the following information in the abstract and added it in the Conclusions Section (IV), showing evidence from the paper.

Each time the gaming agent plays this simple game, it takes on average around 100 steps to reach the target spot. We trained a classifier to recognize Error Potential from observational human critics that watch the agent playing the game. We let the agent play again and we mark movements that trigger an error potential from the human critic. We used those movements as rewards in a Reinforcement Learning scheme, and use them to train a Q-Table. When we let the agent plays the game again, the number of steps that requires to solve the game is now reduced. If we provide feedback based on random signals, we verified that no reduction is achieved and the average number of steps does not change. This shows that there is an effective transfer of information from the brainwaves to the agent. As this process is repeated, the agent keeps improving solving the game effectively, i.e. performing the minimum number of required steps to reach the goal.

Line 40 “This information is used to make a gaming agent <that> improves ....”

This text is no longer contained in the manuscript. We verified throughout the Manuscript to avoid similar mistakes.

Line 11 col 2 “how biological agents learn from its <their> environment by exploring it and getting feedback rewards, either negative or positive.”

We fixed this grammatical error. Thank you very much for noticing it.

Line 17 col2 ”~~Nonneglected~~ is the influence of DeepBrain’s AlphaGo project,” // Re-word

This was already solved in a previous version, and the manuscript no longer contains this line. We appreciate a lot for your suggestion.

Line 24 col 2 “The papers [8], [9], [10] have successfully demonstrated that a robot can be controlled by obtaining a reward signal from a person’s brain activity, ~~which~~ who is observing the robot, <to> solve a task.”

Fixed issue.

Line 58 Put in a reference number from ethics committee from which approval was sough (e.g. University committee)

We added the missing information on Section II.A. Thanks for pointing out this important issue.

Page 2 Line 56 “Data ~~is~~ <are> handled and processed with the OpenVibe Designer,” // datum-data

Fixed issue for every appearances of **data**.

Line 37 col 2 “This dataset has been published on the IEEE DataPort initiative [15].” // I would move this to the end of next paragraph (or possibly end of paper)

Added information in the form of a footnote on Section II.B.

Page 4 Lin 24 This allows ~~to learn~~ the Q-Table <to learn> based on the subject’s feedback from the movements the agent took, which are chosen pseudo-randomly, while executing the brainwave session.

Fixed issue. We appreciate a lot the level of detail in the comments.

Line 50 “The best overall performance is obtained using Logistic Regression.” // also need to discuss the significance of the overall levels of accuracy in Discussion. Is a best performance of 0.672 useful/acceptable? How does it compare to other researchers (if a comparison is possible).

On Section IV, third paragraph we tackled this issue. The overall performance of 0.672 is low. Although a straightforward comparison could be misleading in the context of the ErrP experiment, we added a brief recount of the values obtained for other similar works. We also emphasized this point in future works at the end of the Conclusions section.

Line 42 col 2 “These results are also consistent with their classification ROC curves, shown ~~on~~ <in> Figures 6 obtained for both subjects, where the area under the curve are close to chance level.” // Why didn’t you show one good ROC curve and one bad one. This would provide the basis for comparison.

Due to previous Reviewer’s requests, we added all the ROC curves on the new Figure 6. We now emphasized on this figure’s caption which ROC curves where showing an effective identification of the ErrP potential and which did not. We added this in the text, highlighting the important finding that we could not provide an enhancement of the agent performance based on the rewards from subjects where the obtained ROC curves were not good (Section III, page 5).

Page 5 Figure 5 “Y axis shows the averaged number of steps, while x axis show the number of experiences used to cumulative train the Q-Table.” // Why does Fig5 E only have 2 sets of data? Some discussion of the (smaller) change from 1 – 2 would be appropriate (e.g. I Discussion) Line 51 “No~~t~~ performance gain is evidenced, the agents learn nothing which implies that the reward information is useless.” // useless is not a good choice of wording –provides no value

We fixed the inappropriate wording, and appreciate the Reviewer for pointing this out.

Figure 5, for Subject 5 (in the first version of the manuscript it was labeled E) has only two sets of testing data because that Participant (OHC) couldn’t complete more sessions. Hence we used two training sessions to calculate the Q-Table and subsequently performed the run session (i.e. the agent performs movements randomly based on the trained QTable for 200 iterations). On the other hand, we hypothesize that the smaller change obtained from 1-2 for Subject 5 (as well as for Subject 6), may be due to the fact that the ROC curves for those Subjects also display very low classification accuracy. This could be for reasons related to BCI-illeteracy issues, or the Participant (OHC) not being concentrated at all during the experiment. This information was explained in Section III, page 5.

Line 55 “It can be seen that the overall performance of the agent improves as long as there are more experiences to be used to train it, regardless if they were generated from the brainwaves classification from different subjects.” // more explanation/discussion needed. Is this an average effect from positive learners?

This is a great question. Traditional transfer learning strategies focus on the procedure to enrich a classifier training it with a vast dataset and using it to generalize to new data from unseen scenarios. While dealing with EEG data, this strategy tends to fail, and other approaches are required. We found in this work, that even when that strategy fails, if the system is coupled with a basic reinforcement learning algorithm, the cumulative rewards from different subjects, can be used to effectively transmit information and improve the agent performance. We agree with the Reviewer that this may be due to the average effect of positive or learners with a higher accuracy, but we think it may be related to the fact that the RL algorithm is quite robust and if we can get a higher specificity, the cumulative usage of data from various OHCs help to overcome this issue and improve the agent performance.

This work aims to state whether ErrP signals could be used to train a gaming agent using reinforcement learning. The collected data show that ErrP signals can in fact be classified and used to train an agent effectively. // Can you link these findings to confirm/challenge other recent research in Err potentials

These results are supported by the literature. As we mention in the Introduction, similar experiments have been performed (R. Chavarriaga, A. Sobolewski, and J. d. R. Millán, ”Errare machinale est: The use of error-related potentials in brain-machine interfaces,” Frontiers in

Neuroscience,2014). Our novel approach is to use the rewards to train a gaming agent. We have added this confirmation in the Conclusion Section IV.

Page 6 Fig 8 – why use data from subject 6 (a non-learner)?

The experiment that we performed to generate the Figure 9 doesn't include data from Subject 5 and 6. This is emphasized on the Figure's caption and on the main text in the Result Section. In the Figure 9, the X axis show the progressive number of gaming agent training matches used to update the Q-Table (in this case they are from different subjects).

Line 42 "However, even though this implies that the agent misses frequently that an action taken is wrong, this is not hindering the overall performance and the agent is still learning" //

Line 20 col 2 Despite that, the rewards generated from different subjects can be used to train the same Q- Table to improve its performance, which may lead to strategies where the overall performance is improved based on the information from different human critics at the same time. //These are key findings, worth more discussion/interpretation and inclusion in abstract.

We included both finding in the abstract and we emphasized them in Section IV.

Reviewer 2 General Comments

Comments to the Author

The paper is improved. The contribution is clearer. Thank you for the reply to comments. I attach some further comments.

The authors have satisfactorily addressed my concerns. I have a minor comment regarding Figure 3. It is unclear what the brown bar labelled GrandAvg represents since there is one for each participant. Please clarify. I suggest having another set of bars (a ninth) for the grand average (average over participants) for each of the algorithms used.

We think we misunderstood previous Reviewer's suggestions. We added the brown bar on the previous Figure 3 (now Figure 4) because we assumed the Reviewer was referring to a classification performed using epoched averaged signal segments. We used 5-segments to obtain an averaged signal with an enhanced signal-to-noise ratio (Wim van Drongelen, 4 - Signal Averaging, Editor(s): Signal Processing for Neuroscientists, Academic Press, 2007, Pages 55-70,) per Participant, and performed the classification of ErrP components based on those averaged signals for each Participant. We used only Logistic Regression as classification algorithm, as it was the one with which we achieved better performance for single signal segments. It is clear now that the brown bars were confusing and we have decided to remove them. Instead, we have now included a new Figure 3, where we plotted the **Grand Average** signals over all the participants for the two experimental conditions. The top subfigure shows the signals for the "move closer to the target" condition, whereas the bottom subfigure represents the grand average for the "move further" condition. On the other hand, Figure 9 shows the results, in terms of the decrease in the number of steps for the agent to reach the goal, of using information to train a classifier with data from one Participant and use it to identify rewards for another participant. This is performed in a single-trial approach from individual segments without any signal averaging. We apologize to the Reviewer for our misunderstanding and ask back for more information if it feels like the misunderstanding persisted.