35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Development of an EEG Based Reinforcement Learning Brain-Computer Interface System for Rehabilitation*

Scott A. Roset, *Student Member IEEE*, Hernán F. González, *Student Member IEEE* and Justin C. Sanchez, *Senior Member IEEE*

Abstract— **Spinal cord injuries (SCI) can cause the loss of communication and control of the extremities. The long-term effects of disconnecting the brain from the body can cause plastic reorganization of the motor system that worsens the impairment. New research in coupling standard rehabilitation with developments in Brain-Computer Interfaces (BCI) could engage the user's brain more actively and lead to better performance gains over time. BCI enabled rehabilitation offers the unique ability to rehabilitate the motor system as a whole, including secondary damage in the motor cortex. BCI based rehabilitation is a challenging problem because it leads to both neurophysiological and performance dynamics. Contending with these changes could be problematic for standard BCIs, since they are based on static decoders. To overcome these challenges, an adaptive EEG BCI system is developed here to facilitate the user throughout rehabilitation. The system is based on actor-critic based reinforcement learning (RL) which uses both motor and error related potentials (ErrPs) in the EEG to respond to the evolving performance of the user. Performance of the BCI system is characterized over multiple sessions.**

I. INTRODUCTION

Each year, more than 10 people per million will incur a spinal cord injury (SCI). Of these injuries, one-third is reported to result in tetraplegia [1]. People living with tetraplegia rank hand function as the ability they would most like to see restored [2]. With decrease use of hand movements, plastic reorganization occurs in the brain worsening the impairment [3]. To alleviate motor disabilities caused by spinal cord injury will require rehabilitation that restores not only bottom up function (extremities) but also top-down control (brain activity) to produce a complete therapy for the motor system as a whole.

One approach to produce a more comprehensive therapy is to augment standard rehabilitation with new developments from the study of Brain Computer Interfaces (BCI). BCI's record brain activity and translate it into actions in the physical world [4]. BCI's do this by decoding electroencephalography (EEG) data with a computer system to determine a user's intent. By engaging the user's brain to actively control extremities during rehabilitation, BCI's combined with rehabilitation could offer the unique ability to rehabilitate the motor system as a whole, including secondary damage in the motor cortex [5].

*This work was supported by DARPA RNR subcontract for project W31P4Q-12-C-0200.

S. A. Roset, H. F. González, and J. C. Sanchez are with the Department of Biomedical Engineering, University of Miami, Coral Gables, FL 33146, USA (phone: 305-284-2330 e-mail: jcsanchez@miami.edu).

Rehabilitation of the motor cortex will likely lead to relatively dramatic and rapid changes in user's motor potentials. These changes could be problematic for the BCI, since motor potentials are the primary inputs being decoded. To decode the changing motor potentials will require an adaptive BCI, one that facilitates user performance throughout rehabilitation. Traditional BCI's use a static decoding algorithm to map the user's brain activity to intended actions. Changes to the decoding algorithm typically require the user to participate in a recording session to retrain the BCI however this is undesirable during rehabilitation when therapy is constantly being delivered in real-time.

Recently our lab has developed new reinforcement based decoders that could assist with rehabilitation because they are adaptive in their design [6]. With an actor-critic based reinforcement learning (RL) BCI, the BCI continuously adapts to the user. When the BCI successfully decodes the user's brain activity, that mapping of brain activity is reinforced. Similarly, when the BCI is unsuccessful in decoding the user's brain activity, that mapping of brain activity is adapted.

In this paper, a new EEG BCI system using reinforcement learning is developed as an experimental test bed for augmenting rehabilitation with a BCI. The reinforcement learning architecture behind the adaptive BCI is presented. This adaptive BCI is used to collect sample data trials from a user. These sample trials are then used to run a simulation of the adaptive BCI over an extended period of time. The results of this simulation are then analyzed in the context of augmenting rehabilitation.

II. METHODOLOGY

A. Experimental Task

Since upper extremity function is a top priority for people living with SCI, testing here focused on the ability to control hand grasp. During the experimental task used for initial benchmarking the preliminary system, the user, one healthy, adult, male with no prior BCI experience, watches a display that shows cues to open or close their hand. This display also provides feedback updated after each cue on whether the BCI's decoding was correct using a bar plot to indicate the BCI's decoding of the user's brain activity, Fig. 1. The number of '+'s and '-'s on the screen show the unthresholded output of the motor potentials classifier.

The experiment consisted of four sessions of 120 trials with a 5-minute break between each session. During the first session, a predetermined sequence of cues for “open” and “close” and feedback of “correct” and “wrong” were presented. The user received predetermined feedback of “wrong” 50% of the time to evoke error related potentials (ErrP) in the user’s EEG [7].

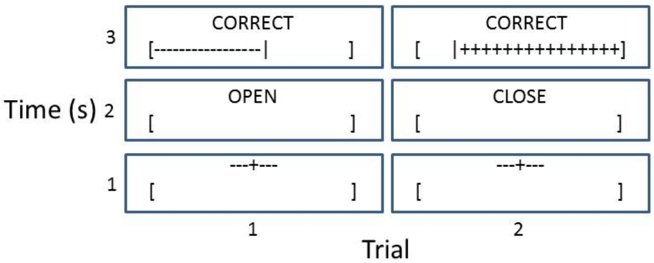

Figure 1. For each trial during the experimental task, the display showed a fixation cross for 1s, followed by a cue for “open” or “close” for 1s, and then feedback of “correct” or “wrong” for 1s. In addition to the explicit feedback of “correct” or “wrong”, a bar plot was presented that show the unthresholded output of the motor potentials decoder.

During the next three sessions, the display produced the decoding of the users modulated motor potentials. Cues of “open” and “close” were presented in a predetermined sequence. And, feedback was displayed, based on the BCI’s classification of the user’s motor potentials.

B. *Actor Critic RL*

When the user controlled the display, the input to the BCI was the user’s motor EEG and the output was two possible actions, “open” or “close”, Fig. 2. The BCI was updated using an actor-critic RL algorithm [8]. The actor-critic RL algorithm tries to optimize the functional mapping of the user’s brain activity to the possible actions. The actor-critic RL method is a semi-supervised machine learning method in which the actor learns from the critic’s feedback [8].

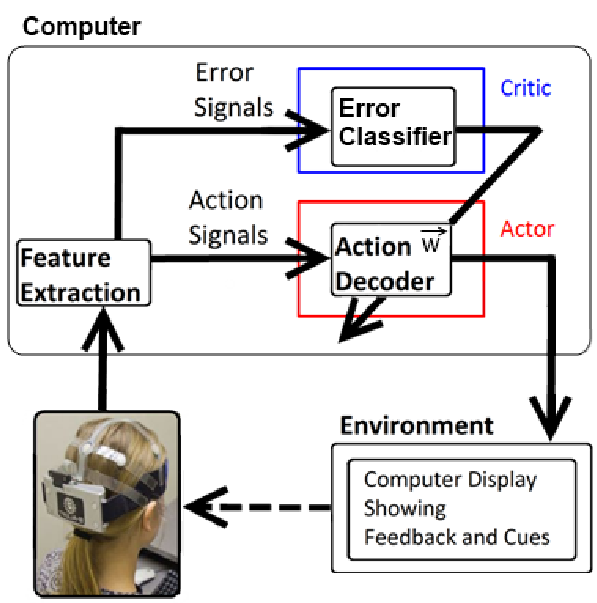

Figure 2. Actor-critic reinforcement learning brain-computer interface architecture. The actor decodes motor potentials and outputs an action shown on the display. The critic detects ErrP and provides feedback to actor. The actor uses feedback from the critic to adapt to the user.

Both the actor and critic are 3-layer fully connected feedforward neural networks. The hidden and output nodes of the neural networks perform a weighted sum on their inputs. The weighted sum at each node is passed through a hyperbolic tangent function with an output in the range of -1 to 1. The weights between the actor’s nodes are initialized randomly and then updated after each trial based on feedback. The critic provides the feedback by decoding the user’s EEG to determine if they generated an ErrP. If an ErrP is detected feedback of -1 is provided to the network for adaptation. If not, a feedback value of 1 is used.

The actor’s weights update can be expressed as:

$$\Delta \boldsymbol{w}_{\boldsymbol{ij}} = \gamma f\left(\boldsymbol{x}_{\boldsymbol{i}}\left(\boldsymbol{p}_{\boldsymbol{j}} - \boldsymbol{x}_{\boldsymbol{j}}\right)\right) + \gamma(1-f)\left(\boldsymbol{x}_{\boldsymbol{i}}\left(\boldsymbol{1} - \boldsymbol{p}_{\boldsymbol{j}} - \boldsymbol{x}_{\boldsymbol{j}}\right)\right) \quad (1)$$

Here w_{ij} is the weight connecting nodes i and j, γ is the learning rate, p_j is a sign function of output x_j (positive values become +1 and negative values become -1) and f is feedback from the critic. The weight update equation is based on Hebbian style learning [9, 10]. Improved classification performance by the actor in early trials was achieved by real time 'epoching' of the data [10]. After each trial, the actor was trained on the current trial and all previous trials.

C. *Data Acquisition*

Neural signals were recorded with a 10-channel Advanced Brain Monitoring (ABM, Carlsbad) wireless EEG system (sample rate 256Hz, 16 bits of resolution) with electrodes in a 10-20 system arrangement. Motor potentials related to the intent to open or close the right hand were collected from the C3 electrode, 1-50Hz. In addition to motor potentials, error-potentials (ErrP) were collected from the Cz electrode, 5-10Hz. EEG corresponding to motor potentials were low pass filtered at 60Hz. ErrPs were low pass filtered at 10Hz [11]. Power spectral density (PSD) of 1 Hz resolution was then computed on the 1s of filtered EEG data after cues were displayed for motor potentials and the 1s of filtered EEG data after feedback was given for ErrPs. The PSD was normalized for each 1 Hz bin by subtracting the mean of all trials and dividing by the standard deviation of all trials.

EEG is commonly contaminated by artifacts originating from ocular muscle motion, which has a high amplitude relative to EEG signals. Ocular artifact such as eye blinks or saccadic motion are relatively simple to identify by visual examination of the neural signals and are characterized by short duration high amplitude wave most present across frontal electrodes (such as Fz, F3, & F4). To remove these artifacts from the signal, Independent component analysis was applied using the Infomax algorithm present in EEGlab [12]. Independent Components due to artifacts from eye movement & blinking were identified by their frontal distribution in scalp topography, matching of component activity in the time domain to eye blink shape, & smoothly decreasing activity power spectrum [13, 14]. The artifactual components were then subtracted from the EEG and the remaining components were remixed to produce a cleaner signal.

D. Critic Error Potential Classifier

The error potential classifier, critic, detects ErrPs in the user's EEG to determine if the user thought an error occurred. The critic than provides binary feedback, -1 or 1, to the actor. The input to the error potential classifier was the PSD from 5-10Hz in 1 Hz bins computed on the 1s of filtered EEG data after the actor's output, action, was shown on the display.

The error potential classifier in the critic is a 3-layer adaptive neural network with 5 inputs nodes and 5 hidden nodes, trained via backpropagation. The 120 trials of the first session were randomly assigned to either a training set or test set. The training set was used to optimize the weights of the critic. The weights produced from the training were assessed by passing the test set through the critic and computing its classification accuracy. The critic was trained and tested until the generalization increased above a threshold [15]. The weights of the critic used in the best testing session were then saved and used for all subsequent experiments.

III. Results

The performance of the EEG BCI based on reinforcement learning is summarized in three parts: 1) critic performance, 2) overall decoding accuracy over time, and 3) characterization of the actor in the early, middle, and late trials of the sessions.

A. Error Potentials Classifier Performance

To test the training paradigm, a 10-fold cross-validation was performed. The above training procedure was repeated 10 times and the average classification accuracy was computed, shown in Table 1.

	ErrP	No-ErrP
Classified as ErrP	69%	37%
Classified as No-ErrP	31%	63%

Table 1. Classification results of the critic for both trials with an ErrP and No-ErrP.

B. Reinforcement Learning BCI Performance

To test the performance of the BCI an offline simulation of 3500 trials was performed. The simulation provided a method to test several factors. A large number of trials could be used, which is more realistic for rehabilitation over several days. Additional processing and filtering could be done on the EEG data, which would require optimization to perform in real time. And, a large number of features could be used as input to the actor, which would also require optimization to perform in real time.

The simulation was performed by generating a random sequence from the 360 recorded trials. The motor potentials from the random trial were filtered and features created, PSD in 1Hz bins from 1-50Hz. Individual frequencies did not show large differences in average power between the two classes (open and closed); however the classifier was able to learn discernible patterns across the 50 frequencies of 1-50Hz. The actor classified the trial based on these features. If the actor's classification was correct for that trial, recorded EEG data from a trial that showed feedback of correct was presented to the critic. Similarly, if the actor's classification was incorrect for that trial, recorded EEG data from a trial that showed feedback of incorrect was presented to the critic. The output of the critic was given as feedback to the actor, so the actor's weights could be adapted with RL.

Fig. 3 shows the cumulative classification accuracy, number of correct trials divided by the number of trials, of the BCI over the course of the simulation. The performance of the BCI increased rapidly over the first few hundred trials and continued to increase until the end of the 3500 trials simulation. After the first 1500 trials, the performance of the BCI showed a monotonic increase, indicating the BCI was converging on a solution and becoming more stable. To test for overfitting the dataset, the same algorithm was run on a surrogate dataset (randomized motor potentials); the end classification accuracy was 51% .

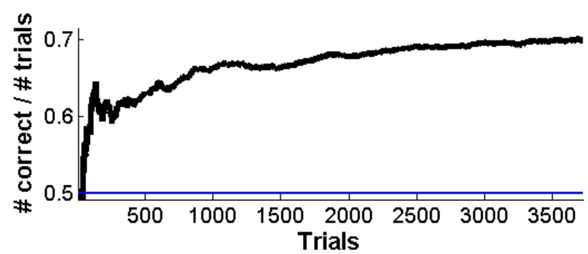

Figure 3. Simulation of a session over 3500 trials, equivalent to a 3 hour recording session. The simulation used 50 features, power in 1Hz bins from 1-50Hz.

Fig. 4 shows a more detailed view of the performance of the BCI during the beginning, middle, and end of the simulation. The actor's performance, the weight values of the actor, and the output of the actor are shown. Again, the BCI performance can be seen to increase rapidly in the beginning of the simulation and become more stable later on in the simulation, while still increasing. The weights of the actor changed dramatically at the beginning of the simulation as the BCI adapts to the user and finds a solution. In the middle of the simulation the BCI is still adapting to the user, as seen in the changing weight values, but is converging on a solution and becoming more stable. At the end of the simulation, as the RL algorithm converged on a solution to mapping the motor potentials to the actions, the weights became stable. The actor's output showed a decrease in errors, red stems, as the simulation progressed. The errors were more likely to be single events and not clustered together, at the end of the simulation.

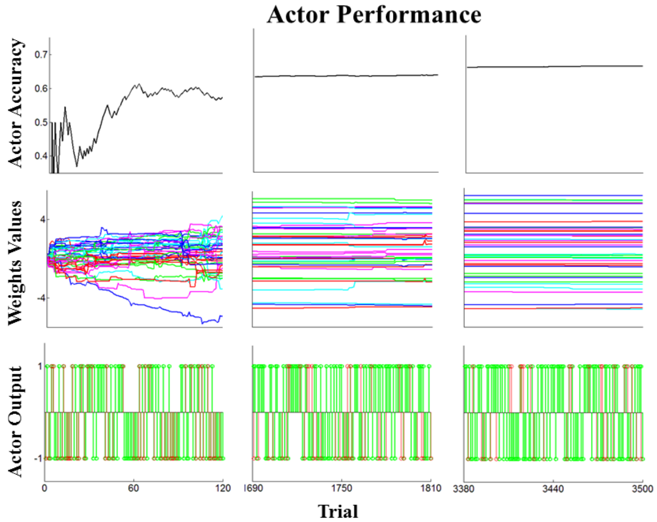

Figure 4. Actor performance during simulation of 3500 trials. Columns show data from different portions of the simulation: beginning, middle, and end. The first row shows the actor's cumulative classification accuracy. The second row shows the actor's weights adapting. The third row shows the actor output, with green stems indicating correct trials and red stems indicating incorrect trials.

IV. Discussion

The simulation showed several results important for rehabilitation. The performance of the BCI increased rapidly in the first few hundred trials. To maintain the user's engagement, the performance of the BCI has to increase above chance quickly, so the user continues to be engaged in control of the device. The performance of the BCI also showed steady increases in later trials, also important for user's engagement. The mapping of motor potentials to actions also became stable in later trials, as seen in the weights values plots. This stability means the user will not see sudden decreases in performance in later trials unless there is a large remapping necessary.

In future work, an EEG system with more electrodes will be used. The additional electrodes will increase spatial resolution within the motor cortex, which could increase the motor decoder accuracy and potentially increase the recognition of ErrPs. When the BCI is paired with rehabilitation (i.e. functional electrical stimulation controlling hand grasp), the subject will see actual physical movement, which should increase engagement in the task. The increased engagement could improve motor potential signal strength. We seek to monitor performance over several days to collect a large number of trials and test how the BCI handles extended breaks between sessions. The extended breaks could lead to more dramatic changes in the user's motor potentials; the adaptive BCI used here is well-suited for this kind of application.

References

[1] M. Wyndaele and J. J. Wyndaele, "Incidence, prevalence and epidemiology of spinal cord injury: what learns a worldwide literature survey?," *Spinal Cord,* vol. 44, pp. 523-529, 2006.

[2] K. D. Anderson, "Targeting Recovery: Priorities of the Spinal Cord-Injured Population," *Journal of Neurotrauma,* vol. 21, pp. 1371-1383, 2004/10/01 2004.

[3] L. R. Hofman and E. C. Field-Fote, "Cortically-Evoked Potentials of Muscles Distal to the Lesion Are Posteriorly Shifted and of Lower Amplitude in Individuals With Tetraplegia Due to Spinal Cord Injury," *Journal of Neurologic Physical Therapy,* vol. 30, pp. 202-203 10.1097/01.NPT.0000281279.62646.04, 2006.

[4] D. J. McFarland and J. R. Wolpaw, "Brain-Computer Interfaces for Communication and Control," *Communications of the ACM,* vol. 54, pp. 60-66, 2011.

[5] J. J. Daly and J. R. Wolpaw, "Brain–computer interfaces in neurological rehabilitation," *Lancet neurology,* vol. 7, p. 1032, 2008.

[6] B. Mahmoudi and J. C. Sanchez, "A Symbiotic Brain-Machine Interface through Value-Based Decision Making," *PLoS ONE,* vol. 6, p. e14760, 2011.

[7] M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein, "ERP components on reaction errors and their functional significance: a tutorial," *Biological Psychology,* vol. 51, pp. 87-107, 2000.

[8] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction* vol. 1: Cambridge Univ Press, 1998.

[9] P. Mazzoni, R. A. Andersen, and M. I. Jordan, "A more biologically plausible learning rule for neural networks," *Proceedings of the National Academy of Sciences,* vol. 88, pp. 4433-4437, May 15, 1991 1991.

[10] E. A. Pohlmeyer, B. Mahmoudi, G. Shijia, N. Prins, and J. C. Sanchez, "Brain-machine interface control of a robot arm using actor-critic reinforcement learning," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2012, pp. 4108-4111.

[11] P. W. Ferrez and J. Millan, "Error-Related EEG Potentials Generated During Simulated Brain-Computer Interaction," *Biomedical Engineering, IEEE Transactions on,* vol. 55, pp. 923-929, 2008.

[12] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods,* vol. 134, pp. 9-21, 2004.

[13] T. P. Jung, C. Humphries, T. W. Lee, S. Makeig, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Extended ICA removes artifacts from electroencephalographic recordings," *Advances in neural information processing systems,* pp. 894-900, 1998.

[14] T. P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects," *Clinical Neurophysiology,* vol. 111, pp. 1745-1758, 2000.

[15] L. Prechelt, "Early Stopping - But When?," in *Neural Networks: Tricks of the Trade*. vol. 1524, G. Orr and K.-R. Müller, Eds., ed: Springer Berlin / Heidelberg, 1998, pp. 553-553.