Reinforcement Learning Brain Machine Interface in Macaques

Marsh BT*, Tarigoppula A*, Chhatbar P, Mahmoudi B, Sanchez J, Francis JT.

Department of Physiology and Pharmacology, SUNY Downstate Medical Center, Brooklyn, NY, 11203

Abstract

The current challenge of brain machine interface (BMI) engineers is to create a BMI that operates successfully in a novel environment. Most systems currently use supervised learning to train their algorithms for a pre-programmed environment. In an endeavor to move beyond supervised learning strategies, our group has created the first reinforcement learning brain machine interface in macaques, where reward expectation is the driving force behind the value of a given action. Using an artificial neural network, we transformed neural vectors from the primary motor cortex into select action-values. Reinforcement learning architecture was used to update the value of eight actions in a four target center-out reaching task for a closed-loop brain computer interface in macaques. We obtained 82% success, if the monkey performed the task manually while the BMI program occurred simultaneously, and 70% success, when the monkey's arm was stationary throughout the task.

Introduction

Neural prostheses have been developed with the goal of restoring quality of life to paralytic patients. Successes to date include the cochlear implant and deep brain stimulators to halt Parkinsonian tremors. Our group has concentrated on subspeciality of neural prostheses involving brain machine interfaces (BMIs). These field encompasses a wide range of technologies and decoding algorithms in an effort to effectively map motor commands to actions. From electroencephalography (EEG) to electrocorticography4 (ECoG) to implantable microelectrode arrays, researchers have been using neurological signals at varying spatiotemporal resolutions as input into their algorithms. Weiner filters2, Kalman filters1, Monte Carlo point-process estimation6, recurrent neural networks5, and actor-critic models3 have been implemented in online control systems. The latter is part of the reinforcement learning literature, which focuses on assigning an expected reward value to actions. The values of these actions are updated throughout a task in a temporal difference (TD) learning paradigm. Consequently, in a changing environment, the reinforcement learning agent can alter its value of a given action in order to accomplish a new goal. With the agenda of creating a BMI that can adapt to real world changes in the environment, we have implemented a TD learning reinforcement learning BMI in macaques for a four-target center-out reaching task.

Methods and Materials

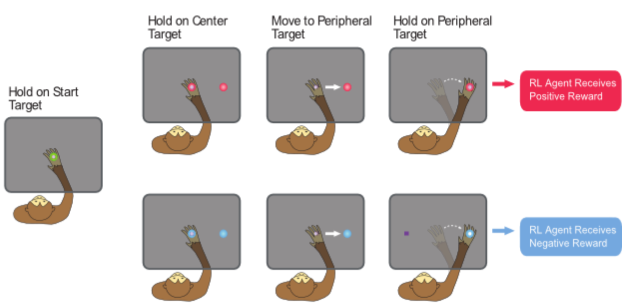


Figure 1. Manual RL Experiment Task. Our subject was trained to perform a center-out delayed reaching task using the Kinarm exoskeleton (BKIN Technologies). The behavioral task consisted of hand movements from a center target to a target located 4 cm to the right of the center target. The target radius was 1 cm. Trials were initiated by entering the center and holding for 300 ms. A liquid reward was provided after a successful reach of the monkey. For the first session of RL experiments, the BMI cursor actions were visible to the subject while it performed the manual task.

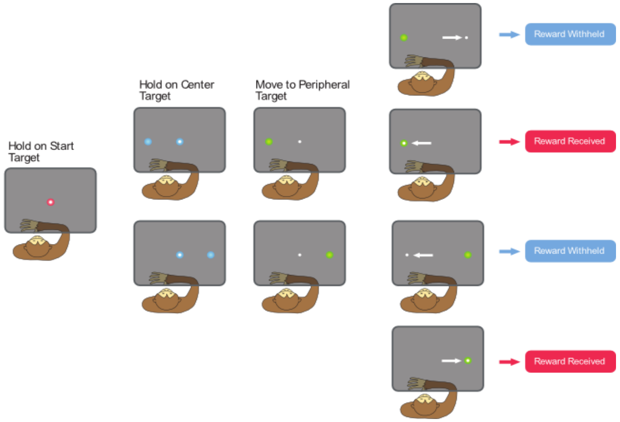
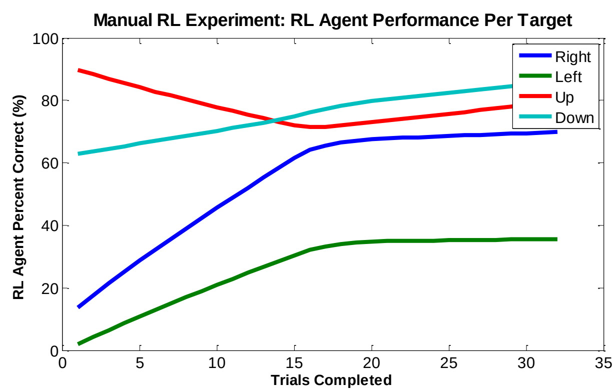
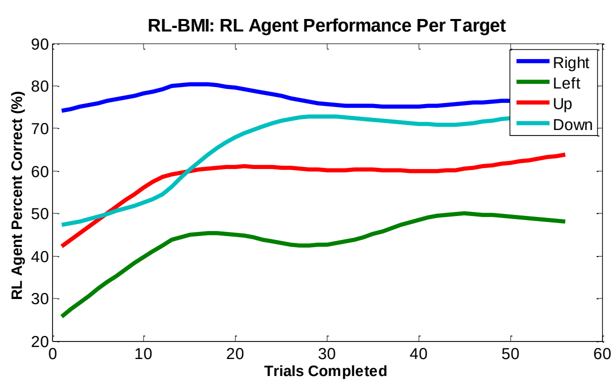
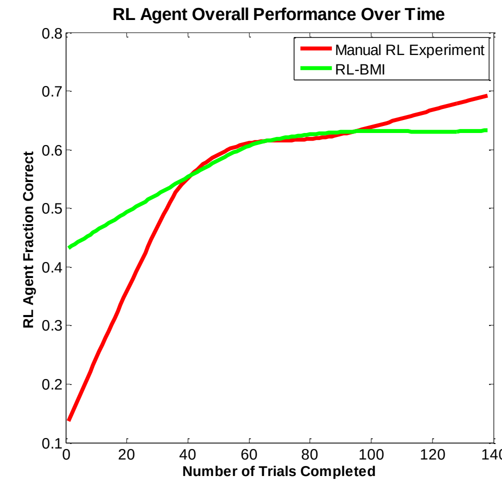


Figure 2. RL-BMI Task. For the second session of RL experiments, the monkey's right arm was kept stationary by securing the Kinarm exoskeleton.







Results

Initially, we ran the reinforcement learning experiment in a manner where the monkey had to manually complete the task, while simultaneously viewing the BMI cursor performance. We achieved 82% success for 4 targets using a one-step TD(λ) algorithm. The four targets were up, down, right, and left, located 4 cm away from the center start target. This higher percent success rate is consistent with the literature, which states that there is higher modulation depth of neurons during a manual task vs. a BMI task. In the next session, the monkey's arm was held stationary during an online RL-BMI experiment. Using a one-step TD(λ) algorithm, we successfully achieved 70% success on the four target BMI.

Discussion

Although we achieved success with the RL-BMI, fifty epochs of offline training was necessary in order to achieve convergence. Thus, assuming continuing success with this algorithm, as we move from the four target case to eight targets, higher computational power will be necessary for convergence. Lower learning rates for the online experiment vs. the offline training files was also necessary to keep performance from dropping. Offline analysis will be required to find the ideal parameters in the eight target paradigm.

References

1. Cunningham JP, Nuyujukian P, Gilja V, Chestek CA, Ryu SI, Shenoy KV. 2011. A closed-loop human simulator for investigating the role of feedback control in brain-machine interfaces. *J Neurophysiol* 105(4):1932-49.
2. Kim SP, Sanchez JC, Rao YN, Erdogmus D, Carmena JM, Lebedev MA, Nicolelis MA, Principe JC. 2006. A comparison of optimal MIMO linear and nonlinear models for brain-machine interfaces. *J Neural Eng* 3(2):145-61.
3. Mahmoudi B, Sanchez JC. A symbiotic brain-machine interface through value-based decision making. 2011. *PLoS One* 6(3):e14760.
4. Shenoy P, Miller KJ, Ojemann JG, Rao RP. 2008. Generalized features for electrocorticographic BCIs. IEEE Trans Biomed Eng. 55(1):273-80.
5. Sussillo D, Nuyujukian P, Fan JM, Kao JC, Stavisky SD, Ryu S, Shenoy K. 2012. A recurrent neural network for closed-loop intracortical brain-machine interface decoders. *J Neural Eng* 9(2):026027.
6. Wang Y, Paiva AR, Príncipe JC, Sanchez JC. 2009. Sequential Monte Carlo point-process estimation of kinematics from neural spiking activity for brain-machine interfaces. *Neural Comput* 10:2894-930.