Correlates of Reward Expectation in the Primary Motor Cortex: Towards Developing an Actor-Critic Model in Macaques for a Brain Computer Interface

Brandi Marsh and Joseph Francis

Abstract:

Reinforcement learning is the branch of machine learning that addresses how an agent should take actions in its environment in order to maximize reward. The Actor-Critic model tackles the reinforcement learning problem by using the agent's brain to create the error signal used for correction of movement in order to maximize reward. Current actor-critic models in reinforcement learning utilize the nucleus accumbens in rodents to decode reward information [1]. These models use the nucleus accumbens as the critic and the primary motor cortex (M1) as the actor in their decoding algorithm. However, in an effort to decrease surgical complications in transformation of this rodent paradigm to the macaque, can information regarding reward expectation be found in higher cortical regions? In specific, does M1 encode a correlate of reward expectation? In other words, can we use M1 as the actor and the critic? Preliminary studies suggest that we can. Using principal component analysis, it seems that we can find a brain state that correlates with expectation of reward in M1.

Introduction:

In beginning efforts to implement the actor-critic model in the macaque for a reinforcement learning brain machine interface (bmi), I have been exploring the possibility of using the primary motor cortex (M1) as a critic for reward expectation. Using simple PSTHs of neural data around the event of reward proved unfruitful. However, transforming the neural data in principal component space resulted in clear state separation of rewarding and non-rewarding trials.

Methods and Results:

The monkey (macaque radiata) sat in an exoskeletal robotic system and made center-out reaching movements to a right target 5 cm away from the start target. The rewarding and non-rewarding trials were color-coded, so the monkey could anticipate reward or no reward. Seven minutes of neural data was placed in 100ms bins and z-scored. Next, principal component analysis was performed on the normalized data. Then, PSTHs of the scores around the event of reward (event 7) and no reward (event 6) were performed for 2 bins before and 2 bins after the event.

With the aim to discover whether expectation of reward could be differentiated in M1 without the monkey moving his arm, an algorithm was created that moved the cursor toward the same target with the same cues of no reward and reward being given to the monkey during the task. In an effort to control for times the monkey was not paying attention, 14 minutes of neural data were analyzed to filter outliers. Although the overall neural signature was different in PCA space from the previous experiment, differentiation between reward and no reward was still possible.

Results for Manual Task

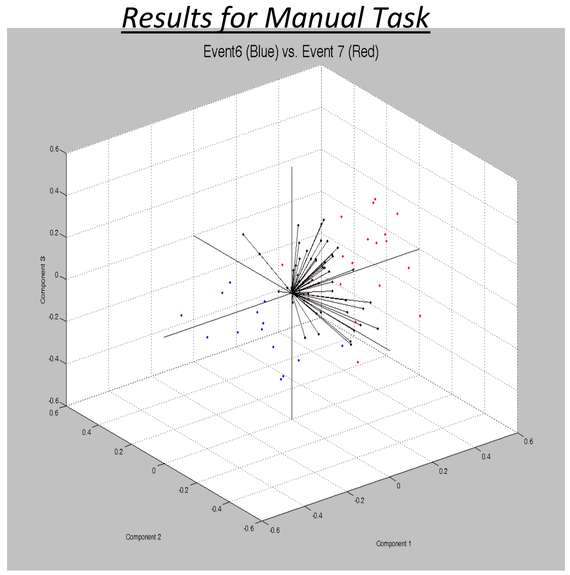
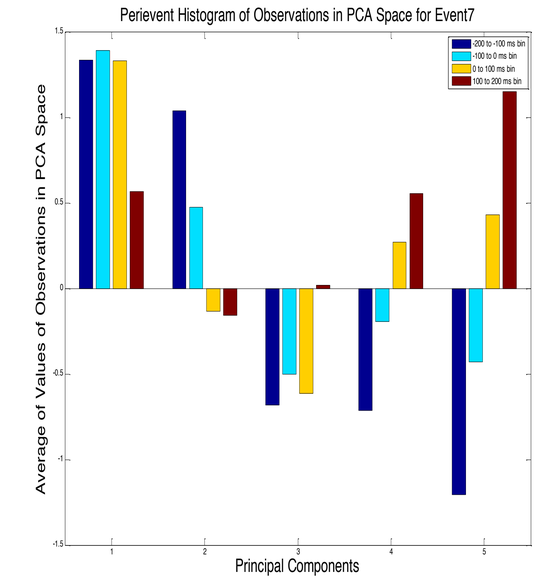
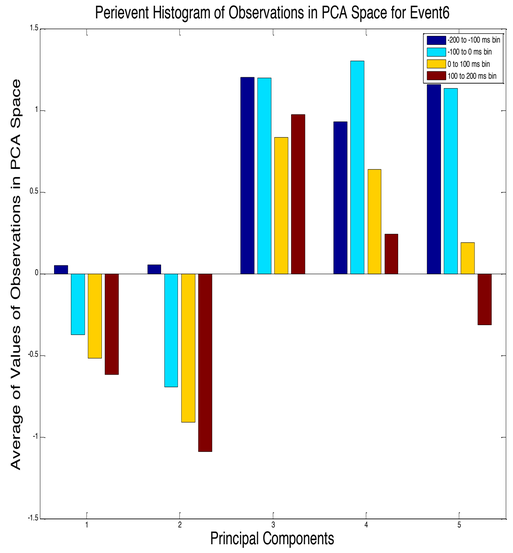


Figure 1: -100 to 0 ms bin in PCA space, using PC 1, 2, and 3.





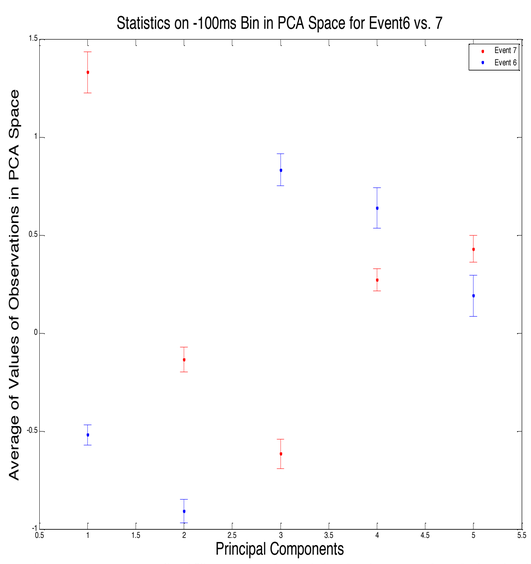


Figure 4: Comparing Averages of Different PCs with their standard deviations.

Results For Passive Viewing

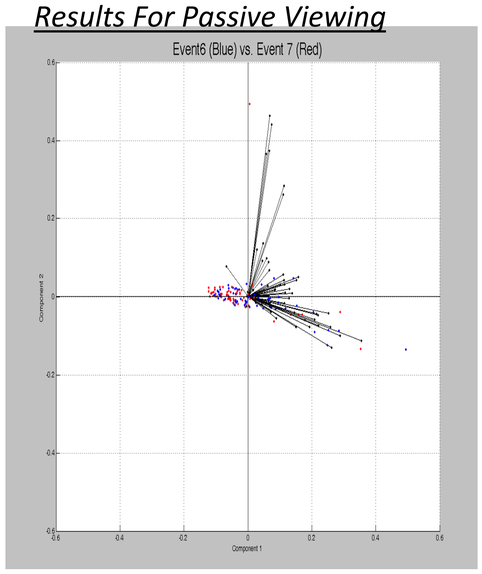
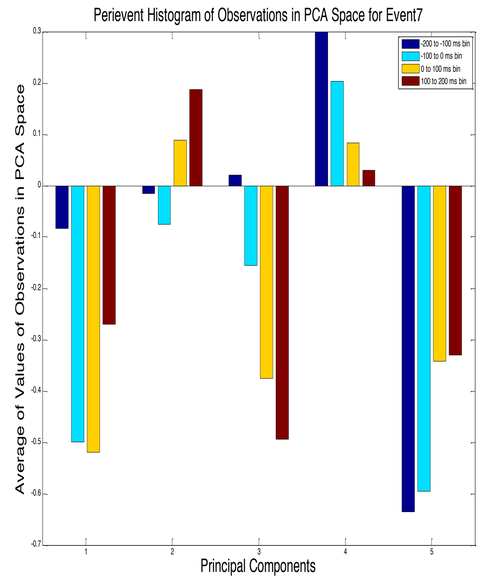
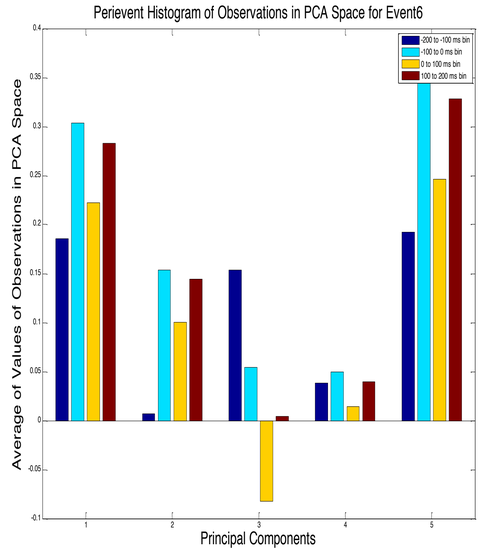


Figure 5: -100 to 0 ms bin in PCA space, using PC 1, 2, and 3.





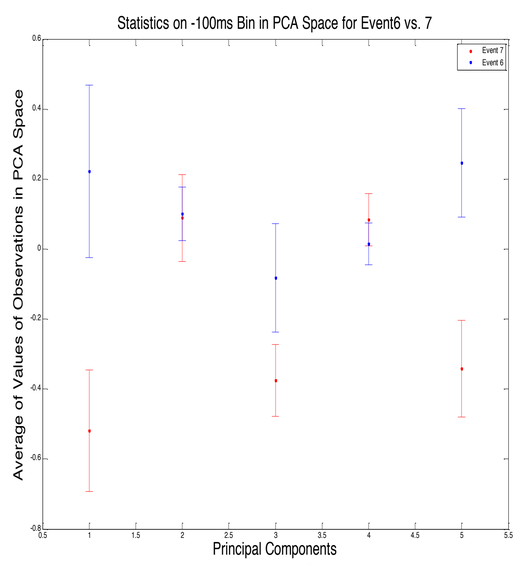


Figure 8: Comparing Averages of Different PCs with their standard deviations.

Discussion:

The next step is to use eye-tracking software to separate points where the monkey is attentive from in-attentive, particularly for the experiment using the algorithm, in order to optimally remove outliers from the data. Additionally, I will compare the kinematics between event 6 and 7 to prove that the observed separation in PCA space is not attributable to differences in kinematics, i.e. I expect the kinematics between the two events to be statistically the same. Lastly, I will develop a classifier to predict the PC separation between rewarded and non-rewarded trials.

References:

Mahmoudi B, Principe JC, Sanchez JC. (2009) An Actor-Critic architecture and simulator for goal-directed Brain-Machine Interfaces. Conf Proc IEEE Eng Med Biol Soc. 2009;2009:3365-8.

William J. Kargo, Botond Szatmary, and Douglas A. Nitz. (2007) Adaptation of Prefrontal Cortical Firing Patterns and Their Fidelity to Changes in Action–Reward Contingencies. J Neuroscience 27(13):3548 –3559.