34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Properties of a temporal difference reinforcement learning brain machine interface driven by a simulated motor cortex

Aditya Tarigoppula*[1], Nick Rotella*[2] and Joseph T. Francis[1]

Abstract: **Our overall goal is to develop a reinforcement learning (RL) based decoder for brain machine interfaces. As an important step in this process, we determine the basic stability and convergence properties of a Temporal Difference (TD) RL architecture being driven by a simulated motor cortex.**

I. INTRODUCTION:

BRAIN-machine interfaces (BMIs) offer tremendous promise as assistive systems for motor-impaired patients [3]. Various supervised decoders with high performance rates have been suggested to map the neural data acquired from the cortex to available actions [8-15]. Reinforcement Learning (RL) based decoders proposed in [2, 3] enable the BMI agent to learn and grow through experience as a natural brain network would [4], without explicit training signals. This computational and biological framework offers a method of neural interfacing that uses goal-directed, experience-based learning to relate neural modulation to behavior through accumulation of rewards and interaction with the environment [4]. Noisy neural tuning curves, measurement noise, loss of neurons, and within-day or day-to-day variations in the neural data are common problems that have to be dealt with appropriately by the decoder, ideally with minimal effect on performance. We performed rigorous simulations to quantify the capabilities of an RL based decoder with respect to these reliability constraints. We present here a series of simulations in which an RL-based decoder was applied to a model of a noisy plastic biological neuronal ensemble.

II. METHODS:

A simulated neuron was developed using the Izhikevich Model [1] as shown below:

$$v' = 0.04 * (v^2) + 5 * v + 140 - u + I \quad (1)$$

$$u' = \eta\,(\beta * v - u) \quad (2)$$

Where, 'v' is the membrane potential of the neuron with a resting membrane potential at 65mV, 'u' is the membrane recovery variable,'I' is the current that goes into the neuron, 'η' and 'β' are dimensionless variables. The parameter 'η' describes the time scale of the recovery variable 'u' [1]. Smaller values result in slower recovery. The parameter 'β' describes the sensitivity of the recovery variable 'u' to the sub-threshold fluctuations of the membrane potential 'v' [1]. A spike was detected every time the membrane potential of the neuron surpassed 30 mV. Given such a model for each neuron, a neural ensemble was developed for our simulations. We started with simple unimodal tuning curves with respect to movement direction and built up assymetric and biomodal tuning curves as follows: Asymmetric curves were created by superimposing two unimodal tuning curves with a peak separation randomly chosen between 30 to 55 degrees [6]. Bimodal curves were created by superimposing two unimodal curves with a peak separation randomly chosen between 125 to 155 degrees [6].

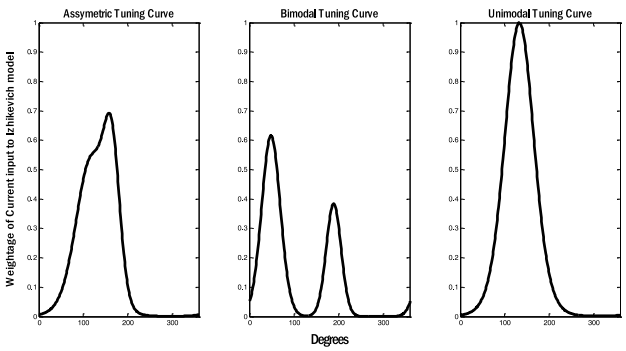

Figure 1: Tuning Curves

The neural ensemble consisted of 80 neurons composed of 60% unimodal, 15% bimodal and 25% asymmetric neurons [6]. The firing rates for these neurons were generated every 100ms to provide a time scale close to firing rates observed during behavior [7]. The tuning depth of our simulated neuron is affected by the ratio of their modulated input current, which is a function of the present movement direction and their tuning curve as seen in Fig.1, to their baseline input current to the Izhikevich model as defined below:

$$I = a * (weightage\ of\ a\ neuron) + b \quad (3)$$

This work was supported by DARPA under grant N66001 awarded to Joseph T. Francis.
* indicates corresponding authors and equal contribution.
E-mail addresses: aditya30887@gmail.com
[1] Department of Physiology and Pharmacology, SUNY Downstate Medical Center, Brooklyn, 450 Clarkson Av, Box# 31, Brooklyn, NY 11203, USA
[2] The Cooper Union, 30 Cooper Square, New York, 10003, USA

Where, $'I'$ is the current that goes into the neuron as shown in Eq (1), 'weighting of each neuron' is the weight obtained from the neuron's tuning curve as shown in Figure (1). By modifying the $'a'$ and $'b'$ parameters of Eq (3), we controlled the baseline and the maximum modulated input current to the Izhikevich model thus effectively controlling baseline and modulated firing rate for each neuron. Henceforth, we will refer to $'a/b'$ of our model as 'Izhikevich-tuning depth' (ITD). Note that "biological noise" was added to the variables $'a'$ and $'b'$ in the form of a white Gaussian noise with standard deviation of 1 to simulate noisy tuning curves.

All the simulations were performed by the RL agent on a center out reaching task wherein, a circular task plane with 4 targets was simulated as shown in Figure (2). The RL agent was provided with only one opportunity to select the correct action out of 8 possible actions to reach the target in a given trial. A trial is considered successful when the RL agent's cursor is on the target and unsuccessful otherwise.

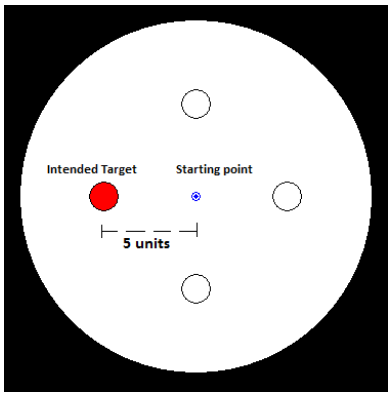

Figure 2: Target Plane for the RL agent. The agent capable of executing 8 possible actions initiates its movement from the starting point and intends to reach the target in one step.

The adaptive nature of the neural ensemble under consideration here was developed as per the conclusions made in [6, 12]. A neuron's preferred direction (PD), which is the direction of movement that causes maximal activity, which lies near a given target is consolidated towards the target at a rate that depends on how far the neuron's PD is at any given time from the intended target. The update in the PD is performed only on the neurons containing PD within 45 degrees on either side of the intended target direction for a given trial. Such a neural consolidation complemented by the improvement of the modulation depth mimics the observed adaptation of the brain while performing BMI (brain machine interface) task. ITD was increased at a rate of 22% per hour [12]. We used the reinforcement learning architecture introduced in [2] and developed further in [3] to control a simulated BMI. We modeled the agent's cursor control problem as a Markov Decision Process (MDP) wherein the RL agent tries to learn the optimal mapping between the neural states and intended action. Specifically, we used Q(λ) learning [5] where Q is called the state-action value function. The firing pattern from this neural ensemble was given as input to a multilayer perceptron (MLP) with one hidden layer consisting 120 hidden units. Update of the weights was performed by back propagation [2, 3]. Eight actions were available to the RL agent. Output units of the MLP state the Q value for each action and the action with the highest Q value is executed 99% of the steps as the optimal action for a given state. Exploratory rate, defined as the percentage of steps in which an action is executed randomly irrespective of its optimality at a given state, was set at 1%. Exploratory rate allows the RL agent to venture out to discover new solutions to the problem, useful especially in an altering environment. The learning rates of the outer layer and inner layer of MLP were 0.005 and 0.01 respectively. Usually, for one simulated session the weights were randomly initialized.

III. RESULTS & DISCUSSIONS:

- Izhikevich-tuning depth vs. RL agent's performance:

Figure (3) presents the performance of the RL agent as a function of the ITD for a neural ensemble consisting of 80 neurons. A tuning depth of 0.75 was found to be needed to consistently yield approximately 95% success. For ITD values above 0.75, the success rate of the RL agent approaches 100% and levels off. This suggests that in selection of a subset of neurons from a recorded population (for dimensionality reduction), neurons with large tuning depths should be preferred.

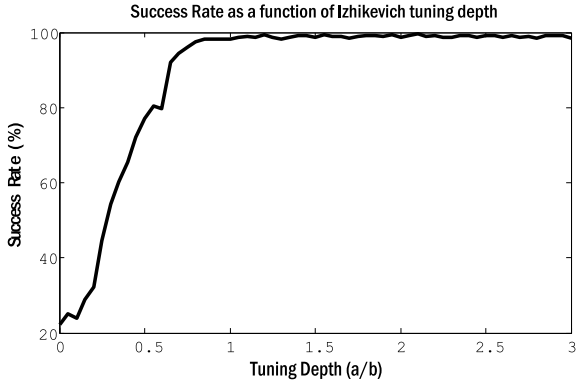

Figure 3: Performance of the RL agent as a function of the Izhikevich-tuning depth for a neural ensemble consisting of 80 neurons

In order to determine the effects of varying other parameters on the adaptive performance of the RL agent, two ITD values were investigated further: a/b = 0.4 and 0.75, which corresponded to a success rate of about 65% and above 95%, respectively.

A stationary RL agent (learning rate reduced to zero; continuing with the weights obtained post convergence) was tested on noisy biological ensemble models with two

different ITDs. The RL agent was found to be capable of performing over 65% when applied to an ensemble with ITD value of 0.75 or more, compared to an ITD value of 0.4, as shown in Figure (4). RL decoding of the ensemble with ITD=0.75 was also more robust to noisier tuning curves.

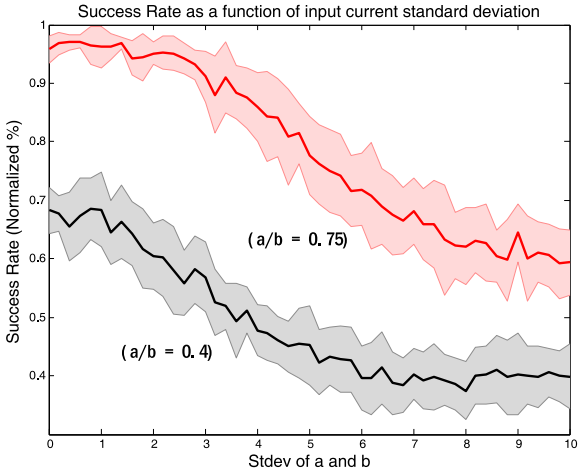

Figure 4: Performance of the stationary RL agent post convergence with respect to "biological noise" for ITD values of 0.4 and 0.75

In addition to biological noise, "measurement noise" was simulated through the addition of white Gaussian noise to the output firing rates generated by the model. Measurement noise simulated the uncertainty introduced in accurately measuring the neural signals due to degradation of electrodes or due to the presence of unwanted external noise.

The magnitude of this noise is specified by its signal-to-noise ratio (SNR), given in dB. Performance of the stationary RL agent was evaluated for ITDs of 0.4 and 0.75 with respect to the measurement noise. The standard deviation of the biological noise for both modulated and baseline currents were set back to 1 for these simulations. Figure (5) shows that with a given ITD, the RL agent required a SNR of about 10dB to perform at its maximum capabilities for each ITD (i.e. a/b = 0.4 and 0.75).

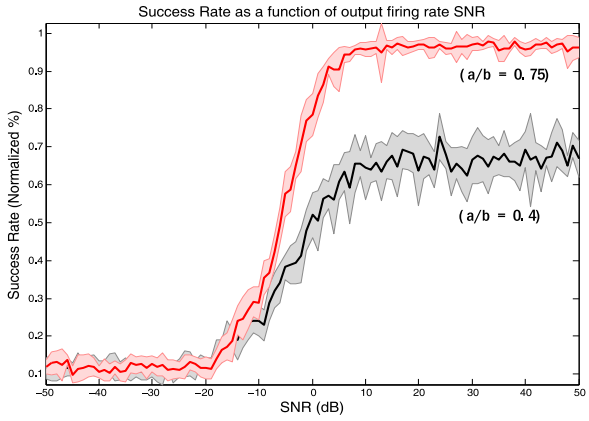

Figure 5: Performance of the stationary RL agent post convergence with respect to "measurement noise" for ITD values of 0.4 and 0.75

- Offline adaptive RL analysis:

As stated in the methods, we co-adapted the neural ensemble wherein the ITD was changed at a rate of 22% per hour [12] whereas the preferred direction of the neurons also consolidated towards 4 target directions over time (a maximum of 90 degree per hour change in preferred direction was allowed). Following convergence, we let the stationary RL agent perform while the neural ensemble continued to adapt at the earlier specified rates. The RL agent performed at 94% success rate (success rate was calculated using only the last 25% of the trials) in a session that lasted for 2000 sec. Please note that the neural ensemble is still adapting throughout the simulation.

- Loss of Neurons:

The size of the neuronal population being recorded may change significantly over the lifetime of an implant and its effect on the BMI decoder's performance is dire [16]. Therefore, we decided to test the RL agent's performance robustness with a continually decreasing neural population. Tuning curves were generated for the 80 neurons as described in the methods section. An ITD (a/b) of 0.75 was employed. The RL decoder was either adaptive (learning rate maintained) or stationary (learning rate reduced to zero) after obtaining convergence on a neural ensemble in order to evaluate and compare the performance of the adaptive system with that of a stationary system when dealing with the loss of neurons. Following an initial simulation of 600 trials utilizing all 80 neurons of the ensemble, the MLP's converged weights were carried on into various simulations, each with a loss of x% neurons per simulation (where, x = 5% or 10% or 15% or...100% of the initial number of neurons in the ensemble for a given simulation). Loss of a neuron meant that its firing rate was set to zero for all 600 trials post convergence. Twenty iterations were performed for every possible value of 'x' in order to average out epoch-to-epoch fluctuations in performance. The order in which inputs were dropped was random. Since the order in which the neurons were dropped could have a significant effect on subsequent performance, this test was performed ten times using ten different randomly-generated orders. Figure (6) displays the average performance over the ten tests. For large neuronal loss, the adaptive RL agent was seen to have little effect on performance over the stationary RL agent apparently because the corresponding neural representation was too limited. Likewise, for very few dropped neurons, the adaptive RL agent does not significantly improve performance over the stationary RL agent because all movement directions were still well-represented. The largest gain in performance by maintaining adaptive nature of the RL agent was achieved around a loss of half of the initial number of neurons. In these cases, enough directional representation remained for the adaptive RL agent to remap its functional relationships from the reduced set of inputs.

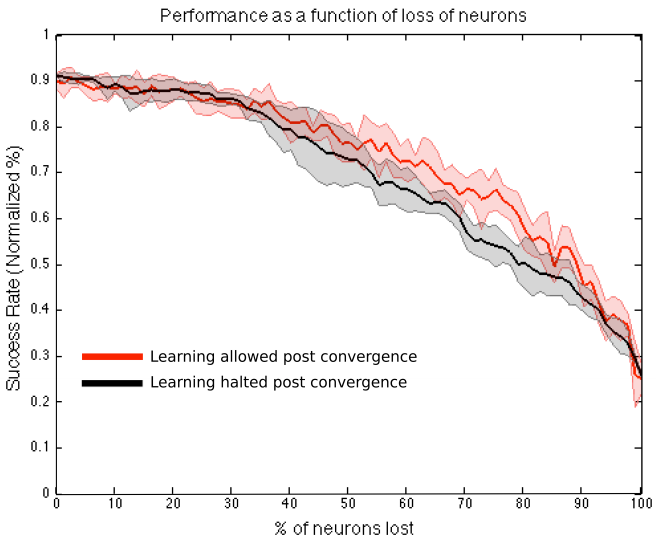

Figure 6: Performance comparison of an adaptive RL agent against a stationary RL agent as a function of loss of neurons

IV. CONCLUSIONS:

For an Izhikevich Tuning depth above 0.75, the RL agent provided us with a performance above 95% for a neural ensemble consisting of 80 neurons. The performance is maintained above 80% even with high biological and measurement noise. The RL agent was also capable of maintaining its performance with almost a 40% loss in neurons for a given level of biological and measurement noise. The simulations suggest that if the neural ensemble adequately represents (modulates) for a given target plane then the RL agent will be capable of obtaining and maintaining convergence from a naïve state. The offline simulations conducted here facilitated a systematic study of the system convergence, parameters, and performance in the presence of varying biological ensemble signal reliability. Now that the foundations have been developed, our goal is to carry out closed-loop experiments engaging the motor and sensory cortices with reinforcement learning.

ACKNOWLEDGMENT

We thank Dr. Justin C. Sanchez and Dr. Babak Mahmoudi for their insight into and facilitation of this project.

REFERENCES:

[1] Eugene M. Izhikevich, "Simple Model of Spiking Neurons." *IEEE Trans. on NEURAL NETWORKS*, VOL. 14, NO. 6, NOVEMBER 2003.

[2] J. DiGiovanna, B. Mahmoudi, J. Fortes, J. C. Principe, and J. C. Sanchez, "Co-adaptive Brain-Machine Interface via Reinforcement Learning." *IEEE Trans. Biomed. Eng.*, 2008.

[3] Justin C. Sanchez, Aditya Tarigoppula, John S. Choi, Brandi T. Marsh, Pratik Y. Chhatbar, Babak Mahmoudi, Joseph T. Francis, "Control of a Center-Out Reaching Task using a Reinforcement Learning Brain-Machine Interface." 5th International IEEE/EMBS conference on Neural Engineering (NER), 2011.

[4] J. C. Sanchez, B. Mahmoudi, J. DiGiovanna, and J. C. Principe, "Exploiting co-adaptation for the design of symbiotic neuroprosthetic assistants," Neural Networks special issue on Goal-Directed Neural Systems, vol. 22, pp. 305-315, 2009.

[5] R. S. Sutton, Andrew G. Barto. "Reinforcement learning: an introduction." The MIT Press, 1998.

[6] Bagrat Amirikian , Apostolos P. Georgopulos, "Directional tuning profiles of motor cortical cells." Neuroscience Research, Volume 36, Issue 1, January 2000, Pages 73–79

[7] Miguel A. L. Nicolelis, "Actions from thought", Nature, 2001.

[8] Chapin, J. K., Markowitz, R. A., Moxon, K. A., and Nicolelis, M. A. L. "Direct real-time control of a robot arm using signals derived from neuronal population recordings in motor cortex." Nature Neuroscience 2, 664-670.

[9] Johan Wessberg, Christopher R. Stambaugh, Jerald D. Kralik, Pamela D. Beck, Mark Laubach, John K. Chapin, Jung Kim, S. James Biggs, Mandayam A. Srinivasan & Miguel A. L. Nicolelis, "Real-time prediction of hand trajectory by ensembles of cortical neurons in primates", *Nature* 408, 361-365, 2000

[10] Mijail D. Serruya, Nicholas G. Hatsopoulos, Liam Paninski, Matthew R. Fellows & John P. Donoghue, "Brain-machine interface: Instant neural control of a movement signal." *Nature* 416, 141-142, 2002.

[11] Shenoy, Krishna V; Meeker, Daniella; Cao, Shiyan; Kureshi, Sohaib A; Pesaran, Bijan; Buneo, Christopher A; Batista, Aaron P; Mitra, Partha P; Burdick, Joel W; Andersen, Richard A, "Neural prosthetic control signals from plan activity" Neuroreport- Volume 14 - Issue 4 - pp 591-596 Motor Systems, 24 March 2003.

[12] J. M. Carmena, M. A. Lebedev, R. E. Crist, J. E. O'Doherty, D. M. Santucci, D. F. Dimitrov, P. G. Patil, and C. S. Henriquez, Miguel A. L. Nicolelis "Learning to control a brain-machine interface for reaching and grasping by primates," PLoS Biology, vol. 1, no. 2, pp. 193-208, Nov. 2003.

[13] W. Wu, M. J. Black, Y. Gao, E. Bienenstock, M. Serruya, and J. P. Donoghue, "Inferring hand motion from multi-cell recordings in motor cortex using a kalman filter," presented at the SAB Workshop on Motor Control in Humans and Robots: On the Interplay of Real Brains and Artificial Devices, Edinburgh, Scotland, 2002

[14] Justin C. Sanchez, Deniz Erdogmus, Jose C. Principe, "A Comparison between Nonlinear Mappings and Linear State Estimation to Model the Relation from Motor Cortical Neuronal Firing to Hand Movements." Proceedings of SAB Workshop on Motor Control in Humans and Robots: On the Interplay of Real Brains and Artificial Devices, 2002.

[15] Justin C. Sanchez, Sung-Phil Kim, Deniz Erdogmus, Yadunandana N. Rao, Jose C. Principe, Johan Wessberg, Miguel Nicolelis, "Input-output mapping performance of linear and nonlinear models for estimating hand trajectories from cortical neuronal firing patterns." 2002.

[16] K. Ganguly and J.M. Carmena, "Emergence of a cortical map for neuroprosthetic control" PLOS Biol., vol. 7, no. 7, p. e1 000 153, Jul. 2009.