Virtual musculoskeletal arm and robotic arm driven by a biomimetic model of sensorimotor cortex with reinforcement learning

Salvador Dura-Bernal
Department of Physiology and Pharmacology
State University of New York Downstate Medical Center
Brooklyn, NY 11203
salvadordura@gmail.com

George L. Chadderdon
Department of Physiology and Pharmacology
State University of New York Downstate Medical Center
Brooklyn, NY 11203
georgec@neurosim.downstate.edu

Samuel A. Neymotin
Department of Physiology and Pharmacology
State University of New York Downstate Medical Center
Brooklyn, NY 11203
samn@neurosim.downstate.edu

Xianlian Zhou
CFD Research Corporation
Huntsville, AL 35805
xxz@cfdrc.com

Andrzej Przekwas
CFD Research Corporation
Huntsville, AL 35805
ajp@cfdrc.com

Joseph T. Francis
Department of Physiology and Pharmacology
State University of New York Downstate Medical Center
Brooklyn, NY 11203
joey199us@gmail.com

William W. Lytton
Department of Physiology and Pharmacology
State University of New York Downstate Medical Center
Brooklyn, NY 11203
billl@neurosim.downstate.edu

Abstract

Neocortical mechanisms of learning sensorimotor control involve a complex series of interactions at multiple levels, from synaptic mechanisms to network connectomics. We developed a model of sensory and motor cortex consisting of several hundred spiking model-neurons. A biomimetic model (BMM) was trained using spike-timing dependent reinforcement learning to drive a simple kinematic two-joint virtual arm in a motor task requiring convergence on a single target. After learning, networks demonstrated retention of behaviorally-relevant memories by utilizing proprioceptive information to perform reach-to-target from multiple starting positions. We utilized the output of this model to drive mirroring motion of a robotic arm.

In order to improve the biological realism of the motor control system, we replaced the simple virtual arm model with a realistic virtual musculoskeletal arm which was interposed between the BMM and the robot arm. The virtual musculoskeletal arm received input from the BMM signaling neural excitation for each muscle. It then fed back realistic proprioceptive information, including muscle fiber length and joint angles, which were employed in the reinforcement learning process. The limb position information was also used to control the robotic arm, leading to more realistic movements. This work explores the use of reinforcement learning in a spiking model of sensorimotor cortex and how this is affected by the bidirectional interaction with the kinematics and dynamic constraints of a realistic musculoskeletal arm model. It also paves the way towards a full closed-loop biomimetic brain-effector system that can be incorporated in a neural decoder for prosthetic control, and used for developing biomimetic learning algorithms for controlling real-time devices. Additionally, utilizing biomimetic neuronal modeling in brain-machine interfaces offers the possibility for finer control of prosthetics, and the ability to better understand the brain.

1 Introduction

Adaptive movements in response to stimuli sensed from the world are a vital biological function. Although arm reaching towards a target is a basic movement, the neocortical mechanisms allowing sensory information to be used in the generation of reaches are enormously complex and difficult to track [18]. Learning brings neuronal and physical dynamics together. In studies of birdsong, it has been demonstrated that reinforcement learning (RL) operates on random babbling [19]. In that setting, initially random movements initiated by motor neocortex may be rewarded or punished via an error signal affecting neuromodulatory control of plasticity via dopamine [9]. In primates, frontal cortex, including primary motor area M1, is innervated by dopaminergic projections from the ventral tegmental area, and recent neurophysiological evidence points to reward modulation of M1 activity [13]. It has been suggested that similar babble/RL mechanisms may play a role in limb target-learning.

The physical world also plays an essential role in learning and behavior [1]. For example, the selection hypothesis emerges from embodiment: the environment can select neuronal dynamics that are suitable for producing desired behaviors through the agency of a limb or other effector [5]. Embodiment can be used to make predictions for how changes will occur during the perception-action-reward-cycle [12], as the embedding of a system in the environment provides adaptation opportunities. Further levels of embedding are achieved by the development of hybrid systems that use *co-adapting* symbiotic relations between brain (the prosthetics-user) and artificial agents (biomimetic systems and smart prosthetic devices) and among these artificial agents.

Computational modeling of biologically-realistic neuronal networks can aid in validating theories of motor learning and predicting how it occurs *in vivo*. It has been shown that biomimetic brain models (BMMs) can replicate many experimental paradigms, such as sensorimotor learning experiments [2], and accurately reproduce physiological properties observed *in vivo*, including firing rates, stimulus-induced modulations and local field potentials [15]. Recently, learning models of spiking neurons using a goal-driven or reinforcement learning signal have been developed [6], many using spike-timing-dependent plasticity.

The present work extends our previous efforts to create a spiking neuronal model of cortical reinforcement learning of arm reaching [2]. In that paper, we demonstrated the feasibility of the dopamine system-inspired value-driven learning algorithm in allowing a one degree-of-freedom (DOF) forearm controller to learn a mapping from proprioceptive state to motor commands needed to direct the virtual hand to the target. Recently, the arm model was extended to 2 DOFs. This makes the task more complex, because the proprioceptive-to-motor command mapping to be learned requires conjunction of the information at the two different joints. The number of synaptic connections which have active plasticity was increased to handle 2-DOF, adding further challenges, as well as flexibility, for the learning method. The output of the 2-DOF model was recently used to drive a robotic arm in real-time via a network interface [4].

Here we have replaced the simple kinematic arm with a detailed model of a 2-DOF musculoskeletal arm, which includes rigid bodies (bones), joints, muscles and tendons. The kinematics and dynamics of the arm are governed by a set of ordinary differential equations (ODEs) that compute the muscles' activation, length and force, as well as the motion of the arm, at a millisecond resolution. The arm model receives an input excitation signal for each muscle, generated by the BMM motor population. It also feeds back realistic proprioception information, including joint angles, which are used to calculate the error signal during reinforcement learning, and muscle lengths, which are encoded by the proprioception neural population and used to generate the appropriate motor commands. This imposes a new set of realistic constraints to the reinforcement learning system, which we are currently analyzing. The output position information of the BMM-driven musculoskeletal arm is also used to control the robotic arm, leading to smoother, primate-like limb movements.

2 Methods

2.1 Virtual musculoskeletal arm

The musculoskeletal model, developed by CFD Research Corporation, includes 8 rigid bodies (ground, thorax, clavicle, scapula, humerus, ulna, radius, and hand), 7 joints and 4 muscle groups: shoulder extensors (posterior deltoid and infraspinatus), shoulder flexors (pectoralis major and anterior deltoid), elbow extensors (triceps), and elbow flexors (biceps and brachialis). This leading to 2 degrees of freedom: arm flexion and elbow flexion. Deltoid, pectoralis, and triceps have 3 branches, and biceps has 2 branches, all of which are required to maintain the model accuracy, despite branches in each muscle being controlled by the same input signal.

Muscles are an extension of Hills muscle model, where the force depends on the current muscle length, velocity and activation [22, 17, 20]. Activation is a delayed response, governed by an ODE, to the neural excitation input, which means activation always tries to follow excitation. Excitation is provided as a normalized value between 0 and 1, and can be interpreted as indirectly measurable by electromyogram (EMG). At every time step (set to 1 ms), given the input excitation to each muscle, the model calculates the muscle activation, fiber and tendon lengths, force, contraction velocity, and the position and velocity of each of the joints.

2.2 Biomimetic model of sensorimotor cortex and reinforcement learning

The entire closed-loop learning system architecture is shown in Fig. 1. A Brain and its interaction with an artificial Environment were modeled, with the Environment containing the Virtual Arm, which was a part of the simulated agent's body, and a Target object which the agent was supposed to reach for. The Virtual Arm possessed two segments (upper-arm and forearm) which could be swiveled through two joints (shoulder and elbow) so that the arm and hand were able to move in a planar space. Each of the arm joints possessed a pair of flexor and extensor muscles for increasing / decreasing the angles, respectively, and which output a muscle length signal to the degree that the muscle was contracted.

One framework for explaining motor reinforcement learning is the perception-action-reward cycle. The learning system is divided into an Actor, mapping perceptions to actions, and a Critic providing reward and punishment feedback to the actor [7]. An Actor system consisting of proprioceptive neurons (P), sensory cells (S), and motor cells (M) was used to control this system. The P cell receptive fields were tuned so that individual cells fired for a narrow range of particular muscle lengths for one of the four muscles. These P cells sent fixed random weights to the S cells, so that the S cells were capable of representing the conjunct of positions in both joints. The S cells, then, sent plastic weights to the M cells, which possessed a separate population of cells for each of the four muscles which was used to generate the muscle excitation. Plasticity was present within the S and M unit populations and between them in both directions. This Actor effectively performed a mapping between limb state, as measured by muscle length, and the muscle excitation for driving each muscle.

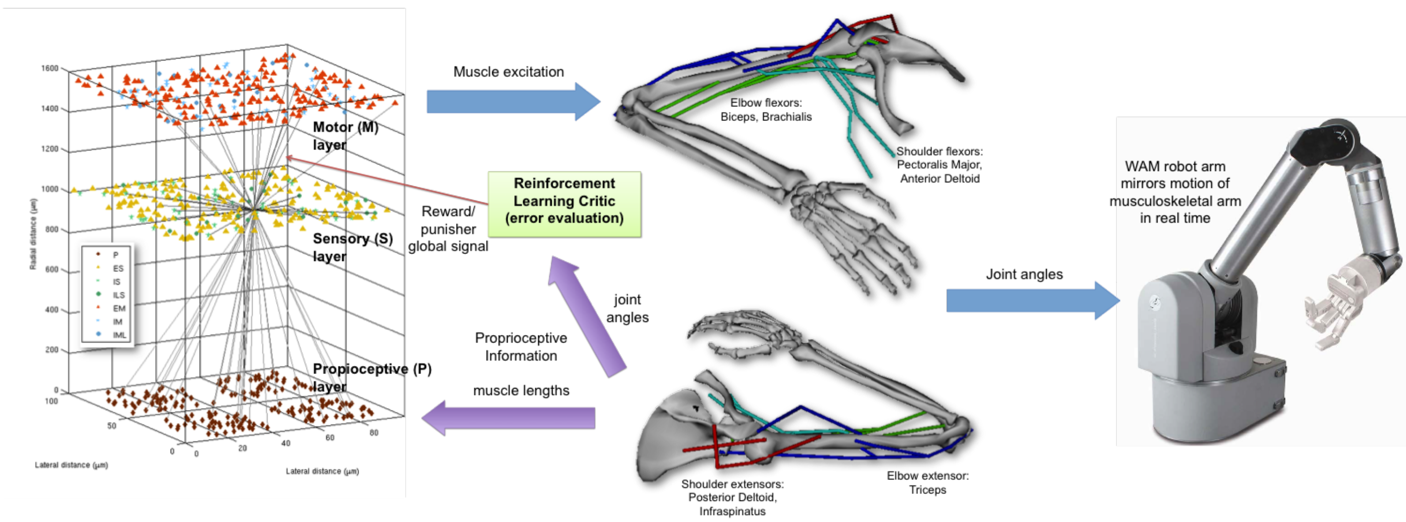

Figure 1: **Closed-loop system implementation with virtual musculoskeletal arm as an intermediate step between the biomimetic model and the robot arm. The virtual arm receives neural excitation from the BMM and feeds back the joint angles, used in reinforcement learning algorithm, and the muscle lengths, used as part of the sensorimotor mapping. The joint angles are also used to drive the WAM robotic arm in real time.**

The Critic component of the system calculated the difference between the hand's location and the target (Error Evaluation), determining from the last two coordinates whether the hand was getting closer or farther away from the target. Based on which was the case, the Critic sent a global reward/punisher signal to the Actor. Plastic synapses kept eligibility traces which allow credit/blame assignment. Rewards caused a global increase in the tagged weights, and punishers caused a decrease, effectively implementing Thorndike's Law of Effect in the system [21], i.e., allowing rewarded behaviors to be "stamped in" and punished behaviors to be "stamped out."

An important component to the system in the Actor was "babbling" noise that was injected into the M cells. An untrained system thus possessed some tendencies to move weakly in a random direction. The Critic, then, was able to allow operant conditioning to shape the motor commands in the context of limb state.

Individual neurons were modeled as rule-based dynamical units with several key features found in real neurons, including adaptation, bursting, depolarization blockade, and voltage-sensitive NMDA conductance [11, 10, 14, 8]. The model consisted of 384 excitatory and 128 inhibitory cells, each with three types of synaptic inputs commonly found in cortex (AMPA, NMDA and $GABA_A$). Cells were connected probabilistically (only a subset of all possible connections were made) with connection densities and initial synaptic weights varying depending on pre- and post-synaptic cell types. The cells were arranged into the three different populations (proprioceptive, sensory and motor) with realistic and anatomical properties. Further details of the model can be found in our previously published paper [2].

3 Results and Discussion

The results in our previous papers demonstrate the flexibility of the network architecture and learning algorithm, which were developed for a 1-DOF virtual arm [2], and later extended to 2-DOFs. The sensory population contained random mappings from the proprioceptive cells, which leads to individual sensory cells forming conjunctive representations of configurations of both joints. The global reinforcement mechanism induced plasticity which shaped the motor population response to the current limb configuration represented in the sensory population. Learning produced alterations in network dynamics, including enhanced neuronal synchrony and enhanced information flow between neuronal populations. After learning, networks retained behaviorally-relevant memories and utilized proprioceptive information to perform reaches to targets from multiple starting positions.

We have now replaced the simple virtual arm with a realistic musculoskeletal arm model. The interface between the BMM and the arm has been implemented using UDP network communication. At every 10 ms time step the BMM sends a packet with excitation values for each of the arm muscles, generated from the motor population spike rate. It then receives from the arm a packet containing the joint angles, used to calculate the error in the reinforcement learning algorithm, and a packet with the muscle lengths, used to form a sensory representation that can be mapped onto the motor commands. Fig. 2 illustrates the position, of the virtual arm, as well as each the muscle group parameters (excitation, activation, length and force) over time. Fig. 3 shows a raster plot of the BMM while interfaced with the arm, illustrating how the proprioceptive population is tracking the four muscle group lengths (compare to Fig.2), and how the motor population is driving the muscle excitation. We have therefore set the ground to commence training and testing the BMM with the new arm. We intend to analyze the network dynamics and plasticity changes that result from using the new realistic arm as compared to using the previous simplistic version.

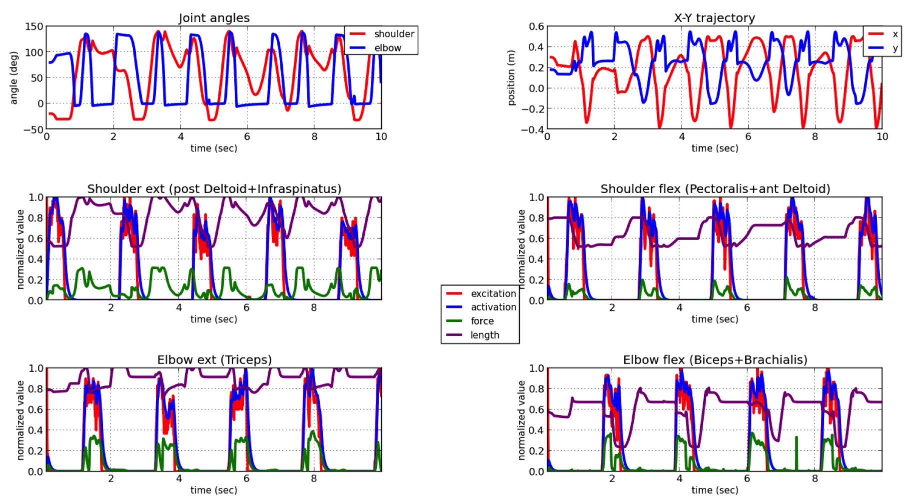

Figure 2: **Joint angles, end-effector position and muscle parameters (excitation, activation, length and force) of the virtual musculoskeletal arm drive by the BMM.**

In previous work [4], we developed a real-time interface between the BMM and a robotic device in the real-world. We used our model to demonstrate the feasibility of using realistic neuronal network models to control devices in real-time. As depicted in Fig. 2, this implementation has been ported to our new system, such that the robot arm will be driven using the joint angles provided by the musculoskeletal arm. This solves two of the main limitations of our previous real-time system[4]: reduces the arm update interval from 50 ms to 10 ms, and provides smoother trajectories, overall leading to more realistic and precise arm reaching movements.

The system presented here is being extended into a framework to link real-time electrophysiological recordings with supercomputers that run the BMMs, and thence to control of prosthetic devices. This co-adaptive brain-machine interface extends the classical BMI paradigm by engaging both subject and computer model in synergistic learning [3, 16].

As our models become more realistic, we will use them as stand-ins for actual brain regions that are damaged or temporarily inactivated. For example, a biomimetic brain model might take input from dorsal premotor cortex (PMd) in subjects that have damage to the motor cortex (M1) and serve to translate the PMd command signals into reaching movements that would have been generated in undamaged M1.

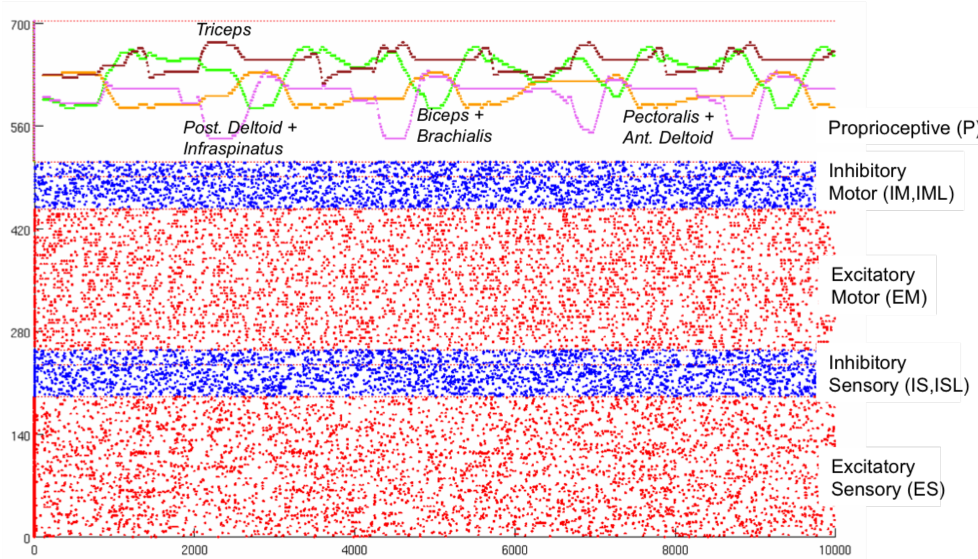

Figure 3: **Raster plot of the BMM, illustrating how the proprioceptive population is tracking the four muscle group lengths (see FIg. 2), and how the motor population is driving the muscle excitation.**

Acknowledgements

Research supported by: DARPA grant N66001-10-C-2008.

References

[1] N. Almassy, G. Edelman, and O. Sporns. Behavioral constraints in the development of neuronal properties: a cortical model embedded in a real-world device. *Cereb Cortex*, 8(4):346–361, 1998.

[2] G. L. Chadderdon, S. A. Neymotin, C. C. Kerr, and W. W. Lytton. Reinforcement learning of targeted movement in a spiking neuronal model of motor cortex. *PLoS ONE*, 7(10):e47251, 10 2012.

[3] J. Digiovanna, P. Rattanatamrong, M. Zhao, B. Mahmoudi, L. Hermer, R. Figueiredo, J. C. Principe, J. Fortes, and J. C. Sanchez. Cyber-workstation for computational neuroscience. *Front Neuroeng*, 2:17, 2010.

[4] S. Dura-Bernal, G. L. Chadderdon, S. A. Neymotin, J. T. Francis, and W. W. Lytton. Towards a real-time interface between a biomimetic model of sensorimotor cortex and a robotic arm. *Pattern Recognition Letters*, (http://dx.doi.org/10.1016/j.patrec.2013.05.019), 2013.

[5] G. Edelman. *Neural Darwinism: The theory of neuronal group selection*. Basic Books New York, New York, 1987.

[6] M. Farries and A. Fairhall. Reinforcement learning with modulated spike timing–dependent synaptic plasticity. 98(6):3648–3665, 2007.

[7] D. Joel, Y. Niv, and E. Ruppin. Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Networks*, 15(4-6):535–547, 2002.

[8] C. Kerr, S. Neymotin, G. Chadderdon, C. Fietkiewicz, J. Francis, and W. Lytton. Electrostimulation as a prosthesis for repair of information flow in a computer model of neocortex. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 20(2):153–160, 2012.

[9] L. Kubikova and L. Kostál. Dopaminergic system in birdsong learning and maintenance. *J Chem Neuroanat*, 39(2):112–123, 2010.

[10] W. Lytton, A. Omurtag, S. Neymotin, and M. Hines. Just-in-time connectivity for large spiking networks. *Neural Comput*, 20(11):2745–2756, 2008a.

[11] W. Lytton and M. Stewart. Rule-based firing for network simulations. *Neurocomputing*, 69(10-12):1160–1164, 2006.

[12] B. Mahmoudi and J. Sanchez. A symbiotic brain-machine interface through value-based decision making. *PloS One*, 6(3):e14760, 2011.

[13] B. Marsh, A. Tarigoppula, and J. Francis. Correlates of reward expectation in the primary motor cortex: Developing an actor-critic model in macaques for a brain computer interface. *Society for Neuroscience Abstracts*, 41, 2011.

[14] S. Neymotin, H. Lee, E. Park, A. Fenton, and W. Lytton. Emergence of physiological oscillation frequencies in a computer model of neocortex. *Front Comput Neurosci*, 5:19, 2011b.

[15] S. A. Neymotin, H. Lee, E. Park, A. A. Fenton, and W. W. Lytton. Emergence of physiological oscillation frequencies in a computer model of neocortex. *Front Comput Neurosci*, 5:19, 2011.

[16] J. Sanchez, W. Lytton, J. Carmena, J. Principe, J. Fortes, R. Barbour, and J. Francis. Dynamically repairing and replacing neural networks: using hybrid computational and biological tools. *IEEE Pulse*, 3(1):57–59, Jan 2012.

[17] L. M. Schutte, M. M. Rodgers, F. Zajac, and R. M. Glaser. Improving the efficacy of electrical stimulation-induced leg cycle ergometry: an analysis based on a dynamic musculoskeletal model. *Rehabilitation Engineering, IEEE Transactions on*, 1(2):109–125, 1993.

[18] R. Shadmehr and S. Wise. *The computational neurobiology of reaching and pointing: a foundation for motor learning*. The MIT press. Cambridge, MA, 2005.

[19] S. Sober and M. Brainard. Adult birdsong is actively maintained by error correction. *Nat Neurosci*, 12(7):927–931, 2009.

[20] D. G. Thelen, F. C. Anderson, S. L. Delp, et al. Generating dynamic simulations of movement using computed muscle control. *Journal of biomechanics*, 36(3):321–328, 2003.

[21] E. Thorndike. *Animal intelligence*. New York: Macmillan, 1911.

[22] F. E. Zajac et al. Muscle and tendon: properties, models, scaling, and application to biomechanics and motor control. *Critical reviews in biomedical engineering*, 17(4):359, 1989.