Research

Article submitted to journal

Subject Areas:
Computational Neuroscience - Machine Learning - Engineering

Keywords:
Biomimetic, Reinforcement Learning, Simulation, Computational Modelling

Author for correspondence:
Raghu Sesha Iyengar
e-mail: bm15resch11003@iith.ac.in

A Novel Paradigm for Deep Reinforcement Learning of Biomimetic Systems

Raghu Sesha Iyengar[1], Kapardi Mallampalli[1] and Mohan Raghavan[1]

[1]Department Of Biomedical Engineering, Indian Institute Of Technology, Hyderabad, INDIA

Mechanisms behind neural control of movement have been an active area of research. Goal-directed movement is a common experimental setup used to understand these mechanisms and relevant neural pathways. On the one hand, optimal feedback control theory is used to model and make quantitative predictions of the coordinated activations of the effectors, such as muscles, joints or limbs. While on the other hand, evidence shows that higher centres such as Basal Ganglia and Cerebellum are involved in activities such as reinforcement learning and error correction. In this paper, we provide a framework to build a digital twin of relevant sections of the human spinal cord using our NEUROiD platform. The digital twin is anatomically and physiologically realistic model of the spinal cord at cellular, spinal networks and system level. We then build a framework to learn the supraspinal activations necessary to perform a simple goal directed movement of the upper limb. The NEUROiD model is interfaced to an Opensim model for all the musculoskeletal simulations. We use Deep Reinforcement Learning to obtain the supraspinal activations necessary to perform the goal directed movement. As per our knowledge, this is the first time an attempt is made to learn the stimulation pattern at the spinal cord level, especially by limiting the observation space to only the afferent feedback received on the Ia, II and Ib fibers. Such a setup results in a biologically realistic constrained environment for learning. Our results show that (1) Reinforcement Learning algorithm converges naturally to the triphasic response observed during goal directed movement (2) Increasing the complexity of the goal gradually helped to accelerate learning (3) Modulation of the afferent inputs were sufficient to execute tasks which were not explicitly learned, but were closely related to the learnt task.

1. Introduction

Emergence of Virtual Physiologies

A digital twin is a simulated model of a physical process of interest. When such a digital twin is connected to real time data from the physical system, it is as if the digital twin is brought to life. Such a digital twin allows continuous monitoring, proactive maintenance and troubleshooting of systems ([1]).This concept when extended to biological systems gives rise to the bio digital twin, also called as the virtual physiological human ([2]) or virtual patient ([3]). Bio digital twin technologies hold the promise of faster development and testing of medical devices and therapies. A key feature of such virtual human models is the ability to reproduce biologically relevant behaviours of interest along with their internal working mechanisms and the associated variability (patient-to-patient and trial-to-trial) These technologies find a natural application in (in-silico) clinical testing(ISCT) ([4]) [1] defined as *'The use of individualised computer simulation in the development or regulatory evaluation of a medicinal product, medical device, or medical intervention'* ([5]). Inspired by the IUPS Physiome initiative[2], the the concept of Virtual Physiological Human (VPH) emerged in 2005[3] and is taking small steps in the direction of improving personalized healthcare ([6]).The commonly modeled virtual physiologies include pharmacokinetic behavior of drugs ([7]),artificial pancreas technology ([8]), cardiac ([9]),motor and musculoskeletal systems ([11–13]).

Interdisciplinary nature of movement modeling

Current models of movement generation are largely musculoskeletal, built in the context of orthopaedics with rudimentary neural activations ([11–13]) or detailed neuronal models with simplified musculoskeleton ([14,15]).This disproportionate focus on one or the other of nervous or musculoskeletal systems arises in large part due to these models being predominantly electrical and mechanical systems respectively. Neuronal ([16–19]) and biomechanical modeling ([20]) being different disciplines in themselves, require different skill sets and methods. Both have a long history of research and development and the state of art in both domains developed independently. In order to create a true state of the art virtual physiology of movement, it is imperative to bridge this chasm by developing methods and composite models that harness the state of art in both disciplines simultaneously. In this work we demonstrate for the first time a class of simple movements using such an exemplar system in the form of a modular, neuro-musculoskeletal upper limb. Built on the NEUROiD platform ([21]) as a co-simulation of the popular open neuronal simulator NEURON ([22]) and musculoskeletal simulator OpenSim ([23]), it allows a choice of model components from both domains in a modular fashion. Such virtual digital models exhibiting movement can play extremely important roles in scientific investigations into movement disorders ([15,17]), development of therapies ([24]) and rehabilitation tools ([25]) for the same.

An integrated stack of motor simulation components generating virtual physiological movement

To virtualize goal directed movement, integration of models of multiple motor subsystems in a hierarchical manner is required. The role of different motor subsystems encompassing the cortex, basal ganglia, cerebellum, spinal cord and musculoskeletal system are well studied ([26]).The purpose of movement is likely to be decided by the prefrontal cortex. Basal ganglia is involved in the choice of motor programs in time ([27]).Cortical Motor neurons activate one or the other spinal motor programs ([28]). Brainstem neurons fine tune the spinal circuits to achieve position

[1] https://www.vph-institute.org/upload/vphinst-position-on-fp8-greenpaper-v3_5192443874603.pdf
[2] http://physiomeproject.org
[3] http://www.vph-institute.org/upload/file517569145f61b.pdf

and balance ([29]).Spinal circuits generate the basic motor programs that coordinate groups of muscles to generate basic movement such as movement in particular directions ([30–33]). These circuits are considered to consist of a layer of rhythm generators and a pattern formation layer that establishes the multiple synergies between muscles([34]). This pattern formation layer could include some of the synergies between the agonist/antagonist muscle pairs. Sensory feedback from muscles and joints is essential to control movements. The muscle spindles are a form of sensory receptors found in the muscle. Ia sensory afferent fibers respond to the rate of change in muscle length, as well as to change in velocity, while the II sensory afferent fibers respond to muscle stretch ([30]). The roles of Ia, II and Ib fibers in motor feedback ([30,31])and the nature of proprioceptive responses during locomotion ([35] has been well studied and modelled ([36]). Anatomy of the spinal cord laminae and the localisation of cell types in them are broadly known ([37,38]). Motoneuron firing results in muscle contraction. In turn the muscle lengths and velocities modulate the activity of the motor neurons and spinal circuits by means of the type Ia, Ib and II afferent fibres ([30,31,36]) forming a neuromusculoskeletal loop. The anatomy of the various muscles and their role in exercising the various degrees of freedom available at the various skeletal joints are well catalogued ([39]).

Many of these motor sub systems have been recreated as mathematical models. Some of the important components are Musculoskeletal models ([20,40]), muscle cell models ([41]),neuro-musculoskeletal reflex circuitry operating in a closed loop ([17,30,31]),spinal cord motor programs([30,31]), pattern generators ([19]), cerebello-thalamocortical models ([42]) and basal ganglia mimicking control strategies to generate complex movements ([14,27, 43]). The integration of these models in a systematic manner that generates a goal directed movement of physiological importance is the holy grail of digital movement twin technology. While significant work has been done in the area of developing such motor components, development of infrastructure for integrated and curated movement infrastructure has been a recent development. Co-simulation platform NEUROiD ([21]) provides a platform that allows integration of state-of-the-art components in musculoskeletal and neuronal domains at various scales in a modular fashion. Building on this work, composite neuro-musculoskeletal upper limb systems have been developed ([44]). In the current work we extend the state of art to develop the equivalent of a central motor program that controls a modular co-simulated neuro-musculoskeletal movement infrastructure. Central motor programs encompass the interneuronal layer of the spinal cord circuitry that establish a variety of muscle synergies along with their cortical control. We also demonstrate how these central programs may be used by spinal cord rhythm generators or by higher cortical centers in order to achieve a goal. However modeling these higher layers themselves will remain out of the purview of the current work.

Control algorithms for moving neuro-musculoskeletal movement systems

Some algorithms use the muscle synergy induced low dimensional control space to control the muscle actuators in order to achieve a given movement ([45]).Design of controllers take advantage of this reduced search space by placing constraints internally (e.g., muscle synergy) or externally (e.g., minimum jerk,torque). Models using Optimal Control Theory find the motor commands by using biologically relevant cost functions such as minimizing the integrated jerk ([46]) or integrated torque ([47]). Sensory feedback was integrated into these models of motor coordination resulting in Optimal Feedback Control of motor actions ([48]). Motor control models have also been developed using neural oscillators ([10]), which demonstrate actions such as walking and running by modifying some of the parameters of the model. Physics-based neuromuscular models ([11]) along with a specific neural circuitry was found to be sufficient to generate steady behaviours such as walking, turning and running. The controller model was further explored for regulating speeds ([12]). A recent study ([49]) also proposes that a model based on bang-bang control (sparse input) is sufficient to observe smooth ballistic hand reaching movements.

Motivation for current work

A common shortcoming of existing methods is that these approaches learn control strategies using observations which are not physiologically observable ([11,12,50]).Most strategies act directly on the musculoskeletal system or do not constrain their action or observation spaces to biophysically accessible variables ([52]). In this work we propose a paradigm for design of the central motor program by means of a Reinforcement learning ([53,54]) based approach. While the controller design is not biophysically realistic, we place constraints on the observation and action spaces to be physiologically relevant. We further constrain the structure of the control network to be compatible with spinal interneuronal networks under the control of cortical signals, thereby making the learnt network functionally and structurally plausible and analogous to a spinal cord interneuron network.

We demonstrate how this learned network is capable of implementing a motor program in the form of patterned supraspinal inputs. We also validate the generated controls by comparing the response from the model against the classic triphasic EMG response ([55,56]). We also demonstrate how this motor program may be used to mimic the execution of repetitive or goal directed movements by rhythm generators in the spinal cord or goal directed movement by higher cortical motor centres.

2. Methods

In the present work we describe the methods for developing a virtual human physiology model that performs a given behavioural task of interest and identify neural signals responsible to achieve the behavioural task. For the purpose of demonstration in this paper, we choose a simple, voluntary, uniplanar, elbow joint movement of the upper limb musculoskeletal model. The goal was set to move the elbow joint to an angle of 0.43 rad $\pm$ 0.05 rad and stay there for a specified duration. In order to achieve this task using a virtual physiological upper limb, the following steps are required:

(a) Building a musculoskeletal model, with requisite anatomy and physiology **(b)** Building a model of spinal cord and relevant circuitry with requisite anatomy and physiology **(c)** Identification of the neuromusculoskeletal interfaces **(d)** Use a learning algorithm to model the supraspinal stimulation patterns, in other words the motor program encoded jointly between the cortex and spinal cord circuits. We do not consider in the present work, models of basal ganglia or cerebellum. We use a modified neuromusculoskeletal model of the upper limb ([44]) implemented using the steps (a)-(c) outlined above. This current work focuses exclusively on the development of a patterned stimulation to the neuro-musculoskeletal upper limb (in steps a-c) that achieves the given task. This system that generates the spatio temporal pattern of stimulation is called as a central motor program (CMP) and is the subject matter of step (d) above. Figure 1 shows the high level architecture used to accomplish the goal.

Building a biologically realistic model of upper limb along with controller circuitry in spinal cord:

In our previous work ([21]), we introduced the NEUROiD platform which supports building large multiscale and multidiscipline neuromotor models. NEUROiD allows micro, meso and macro level definitions of the anatomical, physiological and biomechanical components of the model. Here, scales micro, meso and macro loosely translate to cellular, network, and system, respectively. For the purpose of demonstration in this paper, we use a slightly modified version of the model from [44] developed on the NEUROiD platform. The musculoskeletal model consists of Biceps (long, short), Triceps (long, lateral and medial), Brachialis muscles and an elbow joint, based on [57] and are run on the Opensim simulator ([23]). On the neural side, the C5 and C6 sections of the spinal cord, along with its circuitry and anatomy are modeled and are simulated on NEURON ([22]). Figure 2 (a) shows all the cell groups and spinal circuitry that

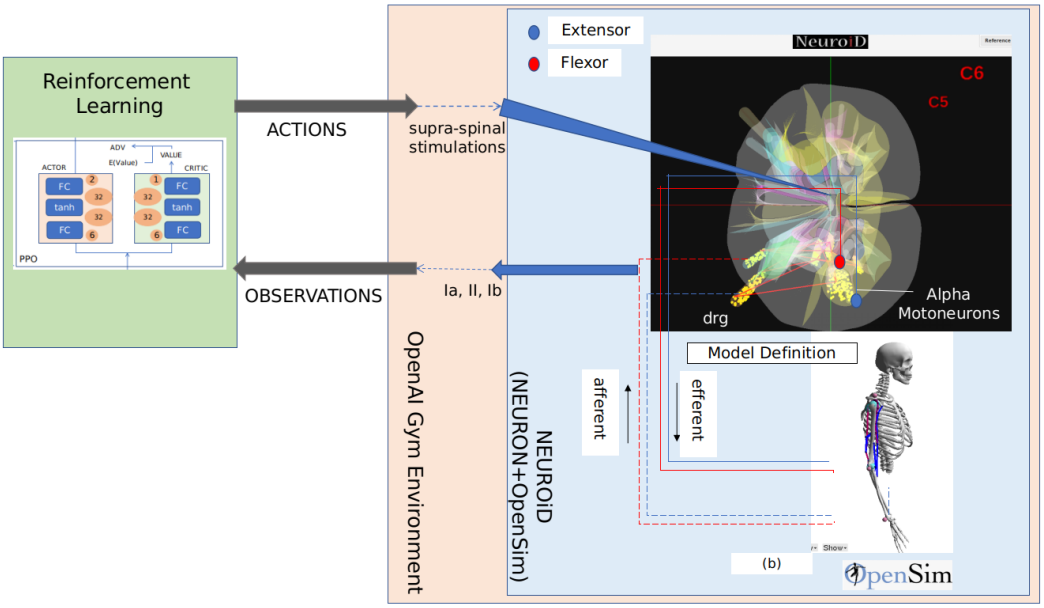


Figure 1. Figure shows the setup used for the work presented in this paper. NEUROiD ([21]) is used to model anatomically realistic C5 and C6 sections of spinal cord with relevant cell groups. NEUROiD uses NEURON ([22]) in the backend for simulation. This model interfaces to an Opensim ([23]) upper body model ([44]). We activate the Biceps long and Triceps long muscles of the Opensim model. The flexor and extensor activations are calculated from the motoneuronal firing of the alpha motoneurons in the C5 and C6 sections respectively. Afferent firings of II, Ia and Ib fibers from the muscles are calculated based on [36] and are used to control the activation of NetStim objects that synapse onto the motoneurons, thus forming the afferent feedback spinal circuit. This NEURON+Opensim setup is wrapped into an OpenAI Gym environment ([58]). The action space in the environment comprises the external stimulation provided to the motoneuron cell groups. This action space is a proxy for the supraspinal inputs received at the spinal cord level during execution of a particular task. The observation space is the II, Ia and Ib afferent firing rates from the muscles. Such an observation space is much closer to biological realism than using the actual joint angle, while modeling the central motor program. We use Reinforcement Learning ([53]) with the Gym environment towards a goal to move the elbow joint to a specific target angle. The rewards in the environment are defined so as to encourage and move the joint towards the target angle and hold its position there.

is present in our example model. We simulated 100 alpha motoneuron cells (Mn), 100 Ia, II and Ib afferent cells each, and 400 interneurons (Iai) in each segment of the spinal cord. The interneurons were modeled as integrate-and-fire cells while the motoneurons were modeled as realistic cell models using a modified Hodgkin-Huxley mechanisms. The efferents consist of the motor neurons innervating the respective muscles in OpenSim. The axonal membrane potentials of the alphamotoneurons in the NEUROiD model were used to calculate the activation values for the corresponding muscles. An action potential from the motoneuron is approximated using $\delta(t)$, which is the unit impulse function. Force produced by motor units is modeled as the output of a linear time invariant system whose unit impulse response is the twitch response ([59]). These activation values for the agonist and antagonist muscles were sent to the OpenSim model. The simulation timestep in NEURON was set to 0.025ms and the muscle activation values were calculated and sent to the OpenSim model every 1ms.

The afferent connections from musculoskeleton into the spinal cord comprise of the monosynaptic Ia and II pathways, Ia and II reciprocal inhibition pathways, Ib excitatory and Ib inhibitory pathways ([30]). The firing rates of the II, Ia and Ib afferent fibers encode the static and dynamic state of the muscle fiber. After a single step of Opensim simulation, the updates in

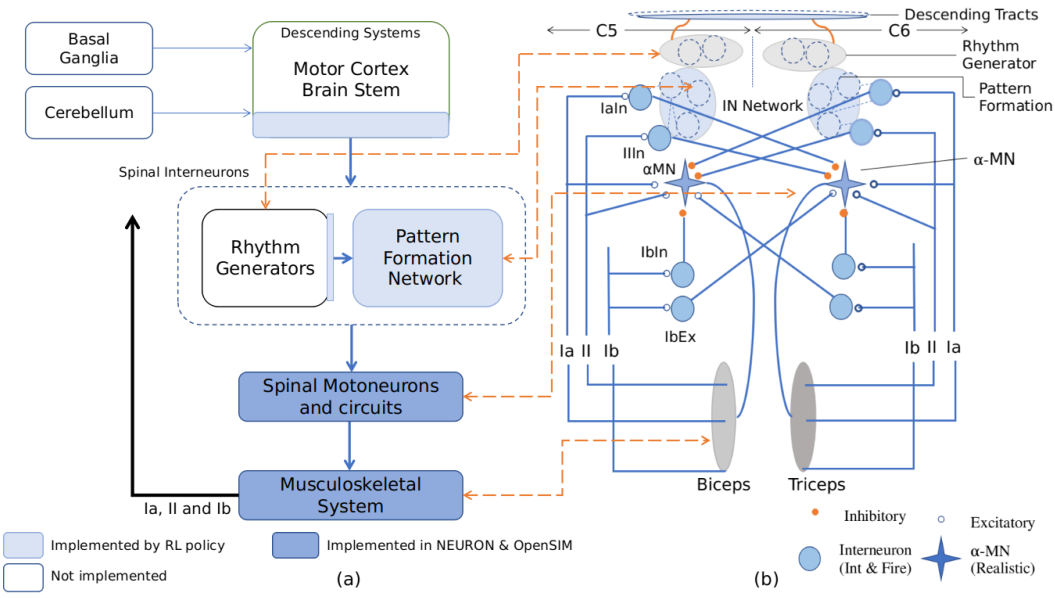


Figure 2. (a) Figure shows the block diagram of a virtual physiology model of the spinal and supraspinal centers involved in movement generation. We defined the spinal circuits in NEURON and used the OpenSim simulator to model the musculoskeletal components. Subfigure (b) shows the cell groups and circuits in more detail. The afferent signals from the musculoskeletal model is fed back to the spinal cord. The reinforcement learning policy implements the Pattern Formation network for the specified task. Interfaces to the cortical structures and Rhythm Generation are also implemented to modulate the activity of the Pattern Formation network. **(b)** Figure shows the cell groups and synaptic connections between them that are modelled in NEUROiD ([21]). We modeled the C5 and C6 segments of the spinal cord. NEUROiD uses NEURON ([22]) in the backend for simulation. Alphamotoneurons are implemented as realistic cell models while the interneurons are integrate-and-fire models. Proprioceptive feedback from the Ia, II and Ib fibers are modelled using the equations from [36]. We modeled monosynaptic Ia and II pathways, Ia and II reciprocal inhibition pathways, Ib excitatory and Ib inhibitory pathways [17]. The interneuron cell groups are also the target of descending connections. The reinforcement learning policy approximates this Pattern Formation (PF) interneuronal network with a deep neural network. The descending connections and the Rhythm Generator (RG) modulate the afferent inputs (section (c)). Equivalent components in the block diagram and the circuit diagram are linked by dotted arrows.

muscle length and velocity were used to calculate the afferent feedback as suggested by [36]. The afferent rates are sent back to the spinal cord simulation running on NEURON, hence controlling the muscles in a closed loop.

The model and simulation parameters are stored in a simulator independent format which can then be used to generate code for execution using various backend simulators. We use NEURON as the backend simulator for the example used in this paper. A 3D rendering of the model with the laminae boundaries, spinal segment outline, cell instances and network connections can be seen in figure 1. We refer the reader to [21] and [44] for further details on the NEUROiD platform and the upper limb model respectively.

Learning a functional equivalent of a central motor program:

The CMPs generate a variety of different movements such as movement of the hand towards the central space, away from the body at various angles in the personal space. The cortical centres and the movement modulators such as basal ganglia utilise the various CMPs in a sequential manner in order to achieve goal directed movements such as reach to grasp. Design of a CMP for a virtual physiology must represent the physiological interfaces, structures and constraints. We impose the

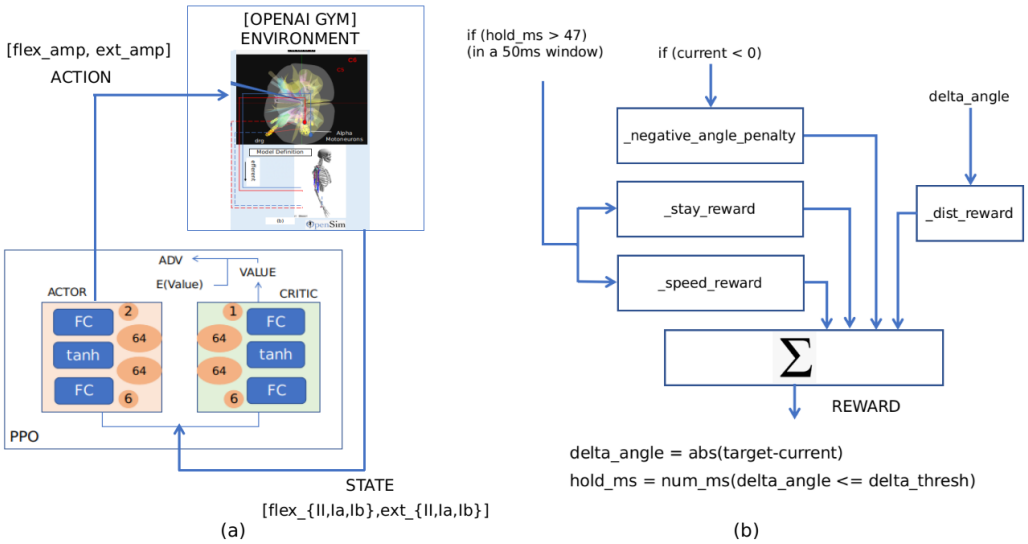


Figure 3. (a) Figure shows the block diagram of a general Reinforcement Learning environment. The agent interacts with the environment by performing action on the environment. The state of the environment and the reward (as a result of the previous action) is read back by the agent. The agent's goal is to identify the optimum policy that determines the set of actions to be taken to maximize the discounted reward accumulated by the agent. The Advantage Actor-Critic network used with the Proximal Policy Optimization ([61]) online policy gradient algorithm that was used to find the optimal policy is also shown. The input state (afferent firings) and output actions (current amplitudes) are shown. Both Actor and Critic are implemented as 2-layered fully connected networks with tanh activation. **(b)** The figure shows the various reward components used in the OpenAI Gym environment. These components are designed to encourage the reinforcement learning policy to identify actions that can move the elbow joint to the target position as soon as possible and hold that position for a sufficient amount of time. *delta_angle* is calculated as the absolute difference between the target angle and current angle of elbow. *hold_ms* is the number of milliseconds for which the angle is held within a small *delta* angle of the target angle. *_dist_reward* is the component of reward that depends on the absolute distance of current angle from the target angle. *_negative_angle_penalty* is the negative reward component to discourage the algorithm from exploring negative angles (which is biologically restricted for the elbow joint). *_stay_reward* is the component of reward that depends on the number of milliseconds for which the elbow has stayed within a small delta angle of the target angle. Finally, *_speed_reward* component of the reward encourages the joint angle to reach the target angle as soon as possible.

following constraints on our CMP under design:

The CMPs execute their actions by stimulation of neurons in their *action space* - the motor neurons and their interneurons. They receive afferent feedbacks from the proprioceptive afferents Ia, Ib and II afferents forming its *observation space*. The design of the functional equivalents of CMPs must comply with these physiologically known observation and action spaces. Further, supra spinal centres modulate the CMPs by means of descending inputs to the interneurons (and motoneurons). Thus the designed CMPs should allow modulation of their behaviour by means of descending inputs. CMPs in physiology are achieved by an unknown network of spinal interneurons (both excitatory and inhibitory) along with some patterned inputs from the primary motor cortex. Thus the designed network equivalent of CMP may use a combination of negative and positive weights. We further impose a condition that the network learnt use as low a number of layers as possible. We start with a single layer network and introduce additional layers as required to achieve the task. The functional equivalent of CMP is designed using a reinforcement

learning ([53]) technique. It may be stressed that although this learning itself is not biological, the resultant equivalent network of CMP is a multilayer neural network with weighted summations and sigmoidal activation functions that are equivalent to biological networks ([60]).

The reinforcement learning algorithm used is the Proximal Policy Optimization ([61]) to find the optimal control policy. Figure 3 (a) shows the PPO setup used for this paper. We used the stable-baselines3[4] implementation of RL algorithms using pytorch. We used a reward discount factor (γ) of 0.99 and λ for generalized advantage estimation was set to 0.95. The clip range (ϵ) for clipping the surrogate function was set to 0.2. Adam optimizer with learning rate of 0.0003 was used for learning. The reward used in the reinforcement learning is composed of multiple components (figure 3 (b)). Training with this reward aids the model to achieve the target position as soon as possible and also hold the position. We used the OpenAI Gym ([58]) toolkit to create a custom environment with a unified interface to encapsulate the NEUROiD model described in sections 2. The *observation space* consists of the II, Ia and Ib firing rates (in Hz) calculated based on the muscle stretch, velocity and force [36]. The stimulation to motoneuron cell groups are implemented as iclamp objects in NEURON, whose amplitude forms the *action space* (refer figure 1). These stimulations are a proxy for the supraspinal inputs and they activate the interneuron and motorneuron network. The equivalent network to CMP is learnt iteratively, with every step involves taking an action, calculating the reward and reading the resultant observations from the environment. In our setup, the NEURON and Opensim solvers are run with step sizes of 0.025ms, 1ms respectively, while the Open AI gym environment is updated every 20ms.

The learning of the CMP equivalent network was attempted with two different configurations - open loop and closed loop. The latter is an upper limb system with the proprioceptive reflexes and the attendant neural circuits intact, while in the former, the proprioceptive afferents back into the spinal cord are disconnected. The openloop training was performed wth tasks of gradually increasing complexity and we used the trained openloop weights to inialize closedloop training. This technique, where a machine learning algorithm is trained with tasks whose complexity is gradually increasing is known as Curriculum Learning ([62]). Such techniques have been applied in robotics and are known to improve speed of convergence and help in generalization. It may be noted that during open loop configuration, the proprioceptive afferents are still available for observation and learning, but the spinal circuits do not have access to them as the reflex loops are open. The modulation is performed by descending fibre-like connections that synapse onto the spinal interneurons. Targets of modulatory signal are the interneurons that relay the afferent signal to the interneuron network (embodied here by the learned network equivalent of CMP). In our scheme the modulatory inputs are of the form of small currents that modify the incoming afferents as seen by the CMP. Modulation of CMP by a spinal rhythm generator is mimicked by a sinusoidal bias current and modulation by cortical centres (for a volitional goal directed movement) by an arbitrarily varying bias (here chosen to be linear, for simplicity).

3. Results

In section (a), we show that the reinforcement learning algorithm was able to learn the necessary policy to achieve a specific goal directed movement, both in openloop and closed loop setup. In section (b), we show that the learnt policy has a close resemblance to Triphasic response ([63]), which is a well known physiological response observed in goal reach tasks. In section (c), we show that modulation of the afferent signals by a small constant can be used to encode the target angle (section i), a potential mechanism to model the supraspinal command to reach different targets. We also show that (section ii) modulation of the afferent signals by a sinusoidal signal can result in a periodic oscillatory movement of the upper limb, a potential mechanism to model central pattern generators.

[4]https://stable-baselines3.readthedocs.io/

(a) Goal reach task with fixed target

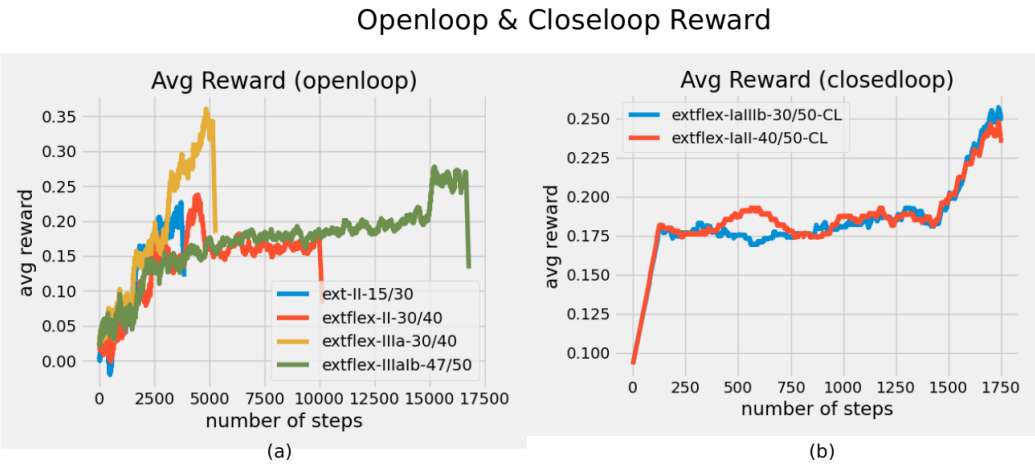


Figure 4. Figure shows the plot of reward Vs the number of steps of reinforcement learning. The reward is averaged over 100 steps. **(a)** This is an open loop setup with the NEURON solver running at 0.025ms, Opensim muscles are activated every 1ms and action space in the Open AI gym environment is updated every 20ms. The plot shows the average reward for slightly different observation spaces and goals. In the above figure, the plots are labelled as follows: **ext-II-15/30** indicates the experiment where we used the afferent II fibers from the extensor as our observation space. We set the goal of learning to move the elbow joint to an angle of 0.43 rad from an initial angle of 0.17 rad. Further, we specify that the angle should be held at 0.43 $\pm$ 0.05 for at least 15 ms in a window of 30 ms. **extflex-II-30/40** indicates the experiment used the afferent II fibers from the extensor and flexor as our observation space. The target was same for all experiments, however we specify that the angle should be held at 0.43 $\pm$ 0.05 for at least 30 ms in a window of 40 ms. The observations are used by the Reinforcement Learning algorithm to calculate the state of next action space. The action space consisted of the current amplitude of the two *iclamp* objects used to stimulate the spinal cord cell groups. **(b)** This is a closed loop setup with the NEURON solver running at 0.025ms. The plots are labelled as follows: **extflex-IIIa-30/50-CL** indicates the experiment where we used the afferent II and Ia fibers from the extensor as our observation space. We set the goal of learning to move the elbow joint to an angle of 0.43 rad from an initial angle of 0.17 rad. Further, we specify that the angle should be held at 0.43 $\pm$ 0.05 for at least 30 ms in a window of 50 ms, in a closed loop setting.

(i) Openloop

Figure 4 shows the plot of average reward Vs number of steps of training in both openlooop and closed loop configurations. We start the training with less stringent goals and increase the complexity gradually.

In the first experiment (in figure 4 (a)), the observation space was restricted to only the II afferents from the extensors. The initial angle of elbow was set to 0.174 rad and the goal of learning was to hold the elbow at 0.43 $\pm$ 0.05 rad for at least 15 ms in a window of 30 ms (ext-II-15/30). We then gradually introduced other afferent firing rates into the observation space and also made the goal more complex. In the next experiment, we add the II afferents flexor also into the observation space and set the goal of learning to hold the elbow at 0.43 $\pm$ 0.05 rad for at least 30 ms in a window of 40 ms (extflex-II-30/40). We observe that the average reward in this case was slightly lesser than the previous case. However, when we introduce the Ia afferents into the observation space keeping the same goal (extflex-IIIa-30/40), we find that the average reward increases and gets better than the first experiment (ext-II-15/30). II and Ia afferents carry the information about muscle stretch and rate of change of muscle length respectively ([30]). Hence, it makes intuitive

Figure 5. Figure shows a typical plot of reward Vs the number of steps of reinforcement learning. A small portion of the plot is zoomed in to show the reward for multiple episodes. We observe that (a) the number of trials for which the reward becomes negative decreases as the training progresses. A negative angle is a biologically invalid state for the elbow joint. Hence, we impose a penalty when the joint tries to move into that state. (b) As the training progresses, the algorithm explores the action space and converges towards actions that result in high reward (>2) for a larger fraction of episodes.

sense that giving diverse information about the system helps in learning. When we further add the Ib afferent firing into the observation space and make the goal even tougher (extflex-IIIAIb-47/50), we find that even in this case, the average reward continues to increase with number of steps of training.

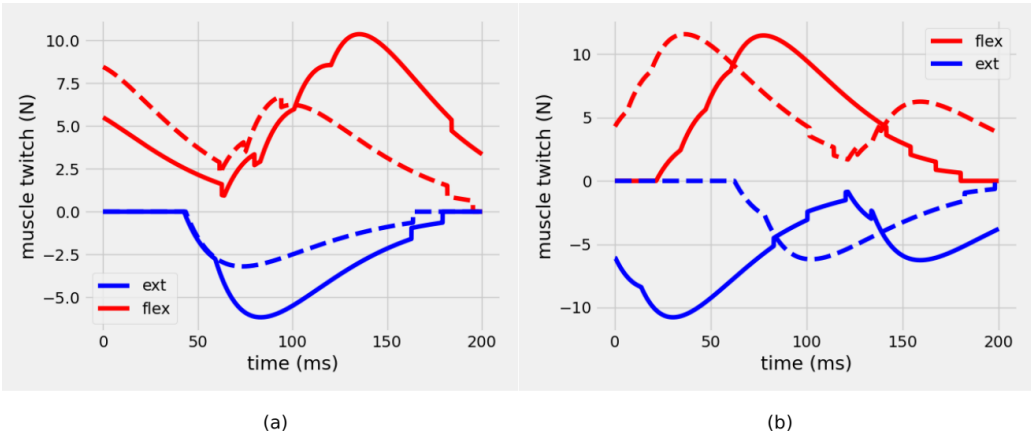
A typical plot of the actual rewards obtained during the training for an openloop experiment is shown in figure 5. The x-axis here represents the number of steps of learning and the y-axis represents the accumulated reward in an episode. The spikes in the plot correspond to the episodes when the reward was high, indicating that the various components of reward (refer figure 3 (b)) were achieved during training.

(ii) Closeloop

Figure 4 (b) shows the plot of average reward Vs number of steps of training in the closed loop setting. We observe that the average reward in the closed loop setting also increases as the training progresses. This improvement in reward was not seen if we performed the close loop training without initialization from the weights trained using open loop. Further, In the closed loop case, we first trained with a relatively easy target as shown in figure 4 (b) (extflex-IaIIIb-30/50-CL) and then used the trained weights from this, to train on a tougher target (extflex-IaIIIb-40/50-CL).

(b) Muscle activation for goal reach task mimics triphasic response

Figure 6 (a) (solid line (S)) shows the twitch of extensor and flexor muscles as observed during the closed loop training after about 12800 steps. Twitch is calculated as the convolution of twitch response of individual muscle and the spike train derived from the axonal membrane potential of the corresponding motoneuron ([59]). This muscle twitch can be considered as a proxy for the EMG signal generated by the muscles during contraction ([59]). Figure 1 from [55] shows the typical triphasic pattern observed in EMG during a goal directed task. We observe that the



(a) (b)

Figure 6. (a) Figure shows the flexor and extensor twitch after 12800 (solid line (S)) and 18900 (dash line (D)) steps of training. The twitch is calculated by convolving the twitch response of a typical muscle with the spike train obtained from the motoneuron firings. We observe that the initial magnitude of flexor twitch in (D) is larger than what was observed in (S). The peak activation of extensor in (D) is also seen to occur slightly earlier compared to what was observed in (S). The peak of the second activation of flexor in (D) occurs much earlier than in (S). All these observations indicate that the algorithm is learning to reach the goal earlier and hold for longer as the training progresses. **(b)** Figure shows the triphasic response observed in the flexor and extensor muscles during training after 18600 steps, when the initial angle is set to 0.7 rad (solid line (S)) and 0.17 rad (dotted line (D)). The target angle is set to 0.43 rad in both cases. We observe that the policy has learnt to activate the flexor and extensor muscles appropriately to reach the goal from both directions.

Reinforcement Learning algorithm has learnt to generate a muscle activation pattern similar to the triphasic response.

Further, when the velocity of such goal directed movement is increased, the amplitude of the bursts is known to increase with a decrease in duration of each burst ([64,65]). Figure 6 (a) also shows the plot of muscle twitch after 18900 (dotted line (D)) steps of training. We observe that as training progresses, the amplitude of initial flexor activation becomes larger and the peak extensor activation occurs earlier for "braking". The second flexor activation for "clamping" is reduced in amplitude as seen in measured physiological response. Since we reward the algorithm if it reaches the goal quickly (refer *_speed_reward* in figure 3 (b)), the algorithm naturally learns this physiologically observed behaviour.

(i) Reach target from an arbitrary starting position

Now, we further train the model by letting the initial angle be a uniform random number between 0.174 rad and 0.697 rad. Figure 6 (b) shows the plot of flexor and extensor muscle twitch with two different initial angles (0.17 and 0.7 rad). We observe that the model has learnt the activation necessary to move the elbow to the specified target angle (0.43 rad) from any initial angle within a range.

(c) Modulation of the CMP equivalent network

As described in section (a) and (b), the reinforcement learning algorithm learns a CMP to generate the stimulation necessary to perform the desired task for a fixed target angle of the elbow. In this section, we describe results of experiments on modulation of this CMP to increase the range and utility of the CMP.

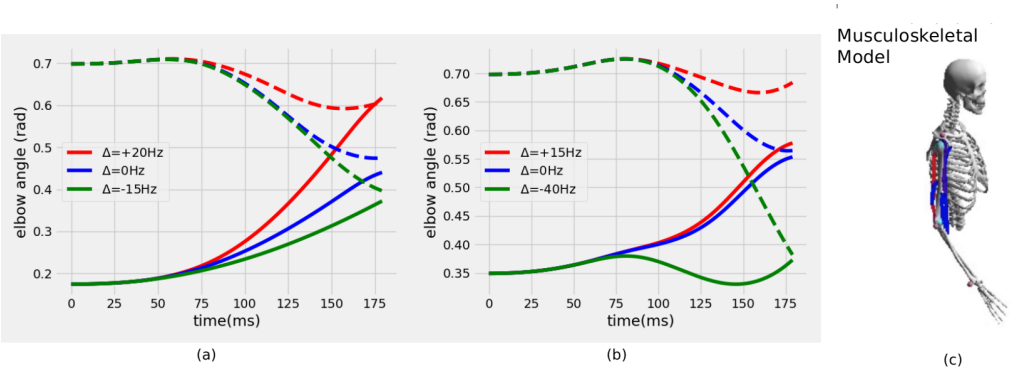


Figure 7. (a) Figure shows the elbow angle for two different initial angle (solid line (S) and dashed line (D)). The afferent signals from the musculoskeletal model were modulated by a constant current signal. Changing the amplitude of the constant signal was found to be equivalent to changing the target angle. We observe that the elbow approached the modified target from both the initial angles. **(a)** is obtined with a simple openloop setup **(b)** is obtained with the closed loop setup with all the cell groups and network connections described in figure 2 (b). **(c)** shows a snapshot of the of the upper limb model ([44]) on OpenSim, which was used for our experiments. We see that elbow is slightly flexed with the antagonist muscle being activated at that point in time.

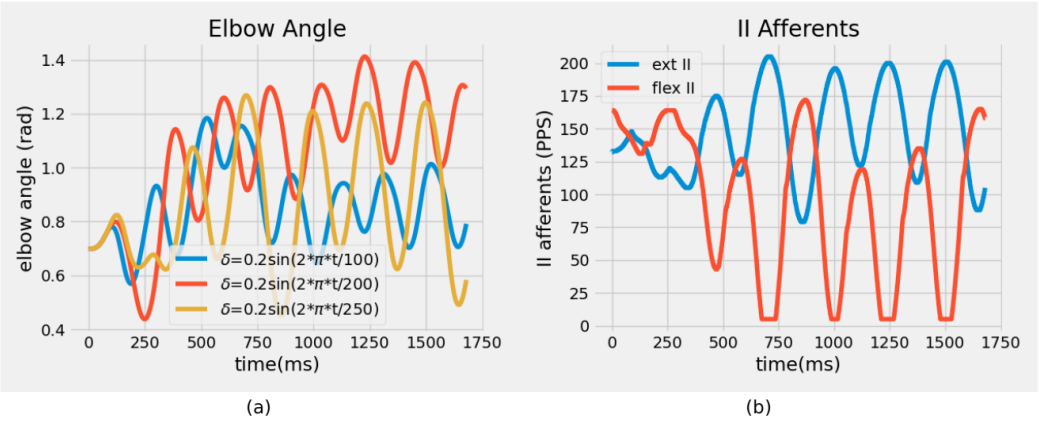


Figure 8. The afferent signals from the musculoskeletal model were perturbed by a sinusoidal current signal. **(a)** shows the plot of the elbow angle when the time period of the sinusoid was set to 100, 200 and 250 ms. **(b)** shows the II afferent signals of both flexor and extensor, when the period of the sinusoidal current was set to 100 ms.

(i) Encodes target (supraspinal trigger)

Figure 7 (a) shows the plots of the elbow angle when a small positive and negative modulatory current are added to the II afferents, thereby modulating their firing rates. The figure shows these plots for two different initial angles, 0.17 rad (solid line (S)) and 0.7 rad (dashed line (D)). We observe that this effects a change in the target angle in a single-joint, uniplanar setup without any further training on the already learnt CMP. Thus, selection of different 'motor programs' to reach different target angles can be achieved by modulation of the descending drive. In this case we were able to execute multiple different motor programs (other target angles) using the same CMP by choosing a target interneuron that effectively altered the observation space as seen by the CMP.

(ii) Generates rhythmic movement (CPG)

A central pattern generator (CPG) is a neuronal network capable of producing rhythmic patterns of motor activity. These could be modified by sensory information and signals from other regions of the central nervous system. Central pattern generators and their role in generation of rhythmic motor patterns ([18]) are studied and also modelled ([19]) in detail.

We modulated the sensory afferent firing rates by a sinusoidal signal and observed that the elbow angle demonstrated a periodic movement, akin to the movement that would be generated by a central pattern generator. We tested with sinusoids having time periods of 100, 200 and 250ms (figure 8 (a)). From the figure, a change in fundamental frequency can be observed when the time period was changed from 100ms to 200ms. Figure 8 (b) shows the II afferent firing rates that was measured from the musculoskeletal model for the case when the time period of the sinusoid was set to 200ms. Thus, we first built a model that learnt only a single motor program to move the elbow to a fixed target angle from a fixed initial angle. We were then able to achieve a periodic movement of the elbow by merely superimposing a small sinusoidal signal on the afferent signals that form the observation space for the reinforcement learning policy.

It is apparent that using any arbitrary form of the modulatory drive, same model can be used to demonstrate an arbitrary uniplanar motion of the elbow joint. Hence, learning a simple motor program under a reinforcement learning setup can be considered to be equivalent to learning a universal motor program provided that the descending modulatory drive targets the appropriate interneurons.

4. Discussion

In this work, we describe a framework to create a central motor program to achieve a specific movement on a virtual physiology consisting of co-simulated neuro-musculoskeletal upper limb. For the purpose of demonstration in this paper, we chose a single-joint, uniplanar movement. However, the techniques used by us are completely generalizable for more complicated movements involving multiple joints and muscle groups. The platform is modular and hierarchical and allows choosing from a variety of musculoskeletal models and spinal neuronal circuit models independent of each other. The model development involves building a composite model of relevant spinal segments along their spinal circuitry (refer section 2) and the musculoskeletal components. Once the observation and action space are identified for the neuromusculoskeletal model, the framework described in this paper (refer section 2) can be used to learn a network equivalent to the CMP. This network delivers the patterned stimulations necessary to achieve the specific movement. The equivalence and physiological plausibility of the network is achieved by constraining the observation space to only the proprioceptive afferents and action space to the motor neurons.

We also demonstrated that using a two layered network, the learned CMP for the task elicited activations resulting in a muscle activation pattern that mimic the physiologically known triphasic response (section (b)) ([32,63]). The first agonist burst is thought to initiate the movement ("action burst"), while the antagonist burst is thought to help in halting the movement ("braking pulse"), and the second agonist burst is thought to dampen the movement ("clamping pulse") ([33]). The temporal scale of this movement (25-150 ms) is consistent with timescales of the movements elicited by motor programs triggered by intracortical microstimulation ([66]). Further, with increasing velocity of movement, the amplitude of the bursts increased consistent with physiological response ([64,65]). Thus our learned CMP may be thought to be playing the role of the pattern forming spinal interneuron circuit as proposed in [67,68]. However, the triphasic muscle response is thought to be due to a central motor program possibly with cortical involvement (([69,70])). Descending connections from primary motor cortex is known to terminate on sets of spinal motor neurons. Further cortical microstimulation of primary motor cortex neurons are known to evoke movements to a fixed target area from a variety of initial positions (([28,71])). These results lead us to hypothesize that our learnt CMP may also be playing

the role of parts of the primary motor cortex and their descending cortical signals along with that of spinal interneuronal circuits.

The task to move the elbow joint towards a specific invariant position from a range of initial positions can be considered as one of the possible motor program that results in a convergence to an invariant final position. Similarly our behavioural repertoire is known to have numerous such manoeuvres as evidenced in daily life and also by intracortical microstimulation studies ([72,73]). However it has been suggested that the number of actual synergies implemented are only a handful ([66]). Our work has demonstrated a possible mechanism by which a rich repertoire of movements can be achieved using a handful of motor programs. In section i, we showed that a small injected synaptic current by way of descending cortical fibres is strategically added to the interneurons that receive the afferents, the effect is a shift in the coordinates of the observation space as seen by the CMP. This results in a shift of the invariant target position. This posits a simple and biologically plausible mechanism by which the same CMP can lead to a family of manoeuvres to different targets. A natural corollary of this simple but powerful result follows when the injected current to the interneurons receiving the afferents is itself sinusoidal. We show that in such a case, movement patterns are cyclical, similar to those elicited under the action of central pattern generators or rhythm generation circuits [67,68,74]. From a generalization of this result it follows that by means of a continuously varying current injection, the target may be continuously varied to trace a trajectory consisting of a sequence of targets.

Our work presents the first instance of biomimetic movement on a co-simulated neuro-musculoskeletal limb. While biomimetic movement has been achieved earlier in custom neuro-muscular models, building virtual movement physiology at scale demands a modular design and the ability to utilize the best advances in the diverse area of neuronal and musculoskeletal modeling. We hope that our work presents a small first step in that direction. We hope to extend our work by realizing the CMP network using models of biological neurons and possibly derive them too using natural learning mechanisms such as STDP, LTP and LTD. Further extensions may be in order to achieve movement to arbitrary target locations in the 3D extrapersonal space. Similarly, adding models for the upstream layers in the form of pre-motor cortices, basal ganglia or cerebellum paves the way for building larger biomimetic systems, one step at a time. Such an *in-silico* movement physiology platform can play a significant role in realizing the benefits in the form of accelerated development of diagnostic, therapeutic, prosthetic medical devices for motor dysfunctions.

Acknowledgment

The authors would like to gratefully acknowledge the support by way of grant-in-aid provided by the Ministry of Electronics and Information Technology(MEITy-4(11)/2018-ITEA), Government of India.

References

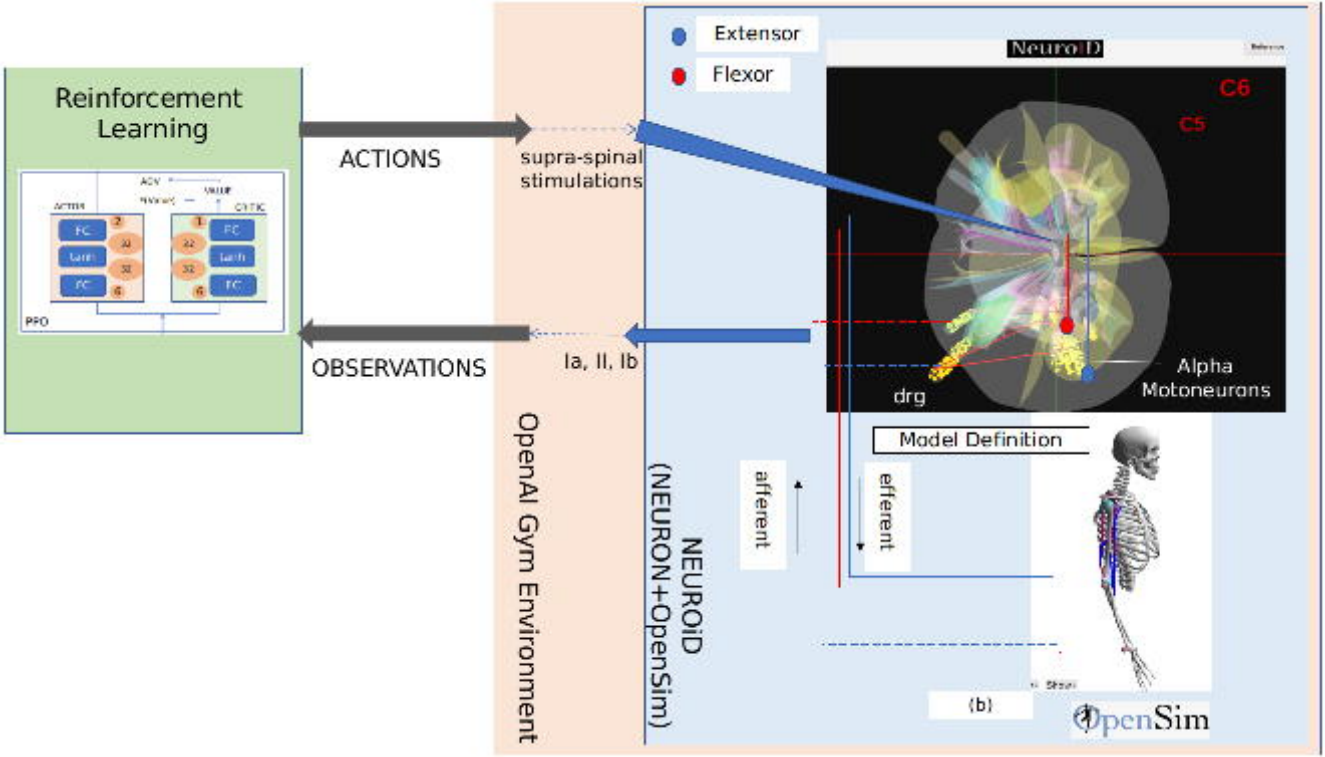
1. Glaessgen E., Stargel D. (2012). "The digital twin paradigm for future NASA and US Air Force vehicles," in Proceedings of the 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA, Orlando: 1818
2. Fenner, J.W et al. (2008). "The EuroPhysiome, STEP and a roadmap for the virtual physiological human". In: Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 366.1878, pp. 2979–2999. ISSN : 1364-503X. DOI : 10.1098/rsta.2008.0089. URL : https://royalsocietypublishing.org/doi/10.1098/rsta.2008.0089.
3. Kononowicz, Andrzej & Woodham, Luke & Edelbring, Samuel & Stathakarou, Natalia & Davies, David & Saxena, Nakul & Tudor Car, Lorainne & Carlstedt-Duke, Jan & Car, Josip & Zary, Nabil. (2019). Virtual Patient Simulations in Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. Journal of medical Internet research. 21. e14676. 10.2196/14676.

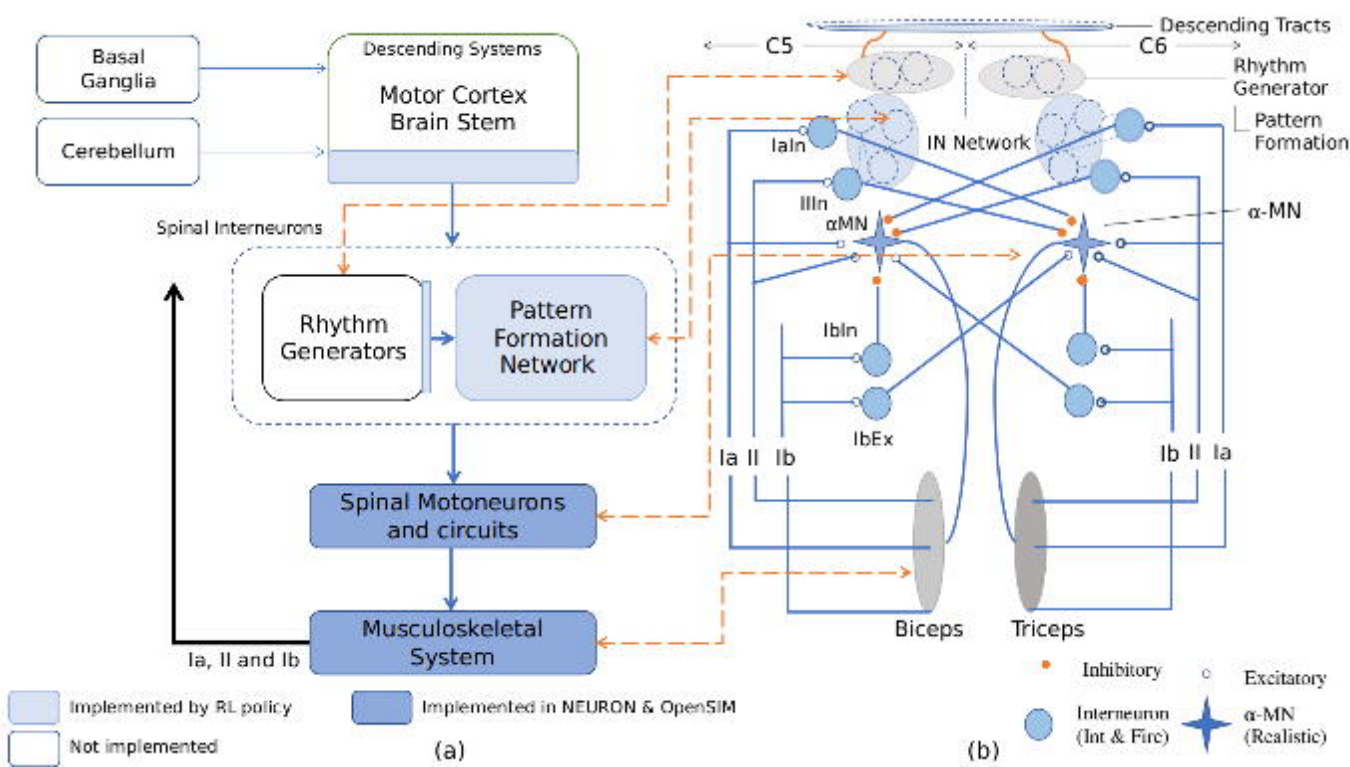
4. Himes, Adam & Haddad, Tarek & Bardot, Dawn. (2016). Augmenting a Clinical Study With Virtual Patient Models: Food and Drug Administration and Industry Collaboration 1. Journal of Medical Devices. 10. 030947. 10.1115/1.4033870.
5. Viceconti, Marco, Adriano Henney, and Edwin Morley-Fletcher (2016). "In silico clinical trials: how computer simulation will transform the biomedical industry". In: International Journal of Clinical Trials 3.2, p. 37. ISSN : 2349-3240. DOI : 10.18203/2349-3259.ijct20161408. URL : http://www.ijclinicaltrials.com
6. Viceconti, Marco, Peter Hunter, and Rod Hose (2015). "Big Data, Big Knowledge: Big Data for Personalized Healthcare". In: IEEE Journal of Biomedical and Health Informatics 19.4, pp. 1209–1215. ISSN : 21682194. DOI : 10.1109/JBHI.2015.2406883.
7. Kumar, Shivani, Suman Chowdhury, and Suresh Kumar (2017). "In silico repurposing of antipsychotic drugs for Alzheimer's disease". In: BMC Neuroscience 18.1.ISSN : 14712202. DOI : 10.1186/s12868-017-0394-8. URL : https://pubmed.ncbi.nlm.nih.gov/29078760/.
8. Kovatchev, Boris (2018). "Automated closed-loop control of diabetes: the artificial pancreas". In: Bioelectronic Medicine 4.1. ISSN : 2332-8886. DOI : 10.1186/s42234-018-0015-6. URL : https://pubmed.ncbi.nlm.nih.gov/32232090/.
9. Baillargeon B, Rebelo N, Fox DD, Taylor RL, Kuhl E. The Living Heart Project: A robust and integrative simulator for human heart function. Eur J Mech A Solids. 2014 Nov;48:38-47. doi: 10.1016/j.euromechsol.2014.04.001. PMID: 25267880; PMCID: PMC4175454.
10. Taga, G. A model of the neuro-musculo-skeletal system for human locomotion II. Real-time adaptability under various constraints. Biol. Cybern. 73, 113–121 (1995).
11. Song, S. & Geyer, H. A neural circuitry that emphasizes spinal feedback generates diverse behaviours of human locomotion. J. Physiol. 593, 3493–3511 (2015).
12. Song, S. & Geyer, H. Regulating speed in a neuromuscular human running model. Proc. IEEE-RAS Int. Conf. Humanoid Robot, pp. 217–222, 2015.
13. Fernandez, Justin & Zhang, Ju & Shim, Vickie & Munro, Jacob & Sartori, Massimo & Besier, Thor & Lloyd, David & Nickerson, David & Hunter, Peter. (2018). Musculoskeletal Modelling and the Physiome Project. 10.1007/978-3-319-58845-2_3.
14. Sreenivasa, Manish, Ko Ayusawa, and Yoshihiko Nakamura (2016). "Modeling and Identification of a Realistic Spiking Neural Network and Musculoskeletal Model of the Human Arm, and an Application to the Stretch Reflex". In: IEEE Transactions on Neural Systems and Rehabilitation Engineering. Vol. 24. 5. Institute of Electrical and Electronics Engineers Inc., pp. 591–602. DOI : 10.1109/TNSRE.2015. 2478858.
15. Moraud, E. M., Capogrosso, M., Formento, E., Wenger, N., DiGiovanna, J., Courtine, G., & Micera, S. (2016). Mechanisms Underlying the Neuromodulation of Spinal Circuits for Correcting Gait and Balance Deficits after Spinal Cord Injury. Neuron, 89(4), 814–828. https://doi.org/10.1016/j.neuron.2016.01.009
16. Capogrosso, Marco et al. (2013). "A computational model for epidural electrical stimulation of spinal sensorimotor circuits". In: Journal of Neuroscience 33.49, pp. 19326–19340. ISSN : 02706474. DOI : 10 . 1523 / JNEUROSCI . 1688 - 13 . 2013. URL : https ://www.jneurosci.org/content/33/49/19326https://www.jneurosci.org/content/33/49/19326.abstract.
17. Pierrot-Deseilligny, Emmanuel and David Burke (2005). The circuitry of the human spinal cord: Its role in motor control and movement disorders. Cambridge University Press, pp. 1–642. ISBN : 9780511545047. DOI : 10 . 1017 / CBO9780511545047.
18. Marder, Eve and Kristina J. Rehm (2005). Development of central pattern generating circuits. DOI : 10.1016/j.conb.2005.01.011.
19. Rybak IA, Shevtzova NA, Lafreniere-Roula M, McCrea DA. 2006. Modelling spinal circuitry involved in locomotion pattern generation: insights from deletions during fictive locomotion. J Physiol (Lond) 577:617–639.
20. Daniel, Matej et al. (2017). Biomechanical analysis of local and global strengthening of gluteus medius. DOI : 10.5606/tftrd.2017.916. URL : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6648248/.
21. Iyengar, R. S., Pithapuram, M. V., Singh, A. K., & Raghavan, M. (2019). Curated Model Development Using NEUROiD: A Web-Based NEUROmotor Integration and Design Platform. Frontiers in Neuroinformatics, 13, 56. https://doi.org/10.3389/fninf.2019.00056
22. Hines, M. L. and N. T. Carnevale (1997). The NEURON Simulation Environment. DOI :10.1162/neco.1997.9.6.1179. URL : https:/pubmed.ncbi.nlm.nih.gov/9248061/.

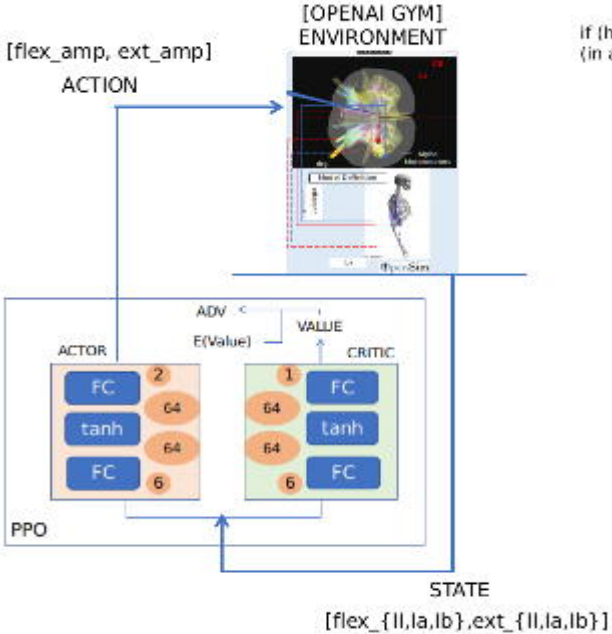
23. Seth, A., Hicks, J. L., Uchida, T. K., Habib, A., Dembia, C. L., Dunne, J. J., Ong, C. F., DeMers, M. S., Rajagopal, A., Millard, M., Hamner, S. R., Arnold, E. M., Yong, J. R., Lakshmikanth, S. K., Sherman, M. A., Ku, J. P., & Delp, S. L. (2018). OpenSim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. PLOS Computational Biology, 14(7), e1006223. https://doi.org/10.1371/journal.pcbi.1006223
24. Cogiamanian, F., Ardolino, G., Vergari, M., Ferrucci, R., Ciocca, M., Scelzo, E., Barbieri, S., & Priori, A. (2012). Transcutaneous Spinal Direct Current Stimulation. Frontiers in Psychiatry, 3, 63. https://doi.org/10.3389/fpsyt.2012.00063
25. Arle, Jeffrey E. et al. (2014). "Mechanism of dorsal column stimulation to treat neuropathic but not nociceptive pain: Analysis with a computational model". In: Neuromodulation 17.7, pp. 642–655. ISSN : 15251403. DOI : 10 . 1111 / ner . 12178. URL : https://pubmed.ncbi.nlm.nih.gov/24750347/.
26. Eric R. Kandel, James H. Schwartz, Thomas M. Jessell, Steven A. Siegelbaum, A. J. Hudspeth (2013). Principles of Neural Science, Fifth Edition | AccessNeurology | McGraw-Hill Medical. URL:https://neurology.mhmedical.com/book.aspx?bookID=1049
27. Gurney K., Prescott T. J., Redgrave P. (2001). A computational model of action selection in the basal ganglia. I. A new functional anatomy. Biol. Cybern. 84, 401–410. 10.1007/PL00007984
28. Georgopoulos, A. P. et al. (1982). "On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex". In: Journal of Neuroscience 2.11, pp. 1527–1537. ISSN : 02706474. DOI : 10.1523/jneurosci. 02-11-01527.1982. URL : https://pubmed.ncbi.nlm.nih.gov/7143039/.
29. Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., & Williams, S. M. (2001). Motor Control Centers in the Brainstem: Upper Motor Neurons That Maintain Balance and Posture. https://www.ncbi.nlm.nih.gov/books/NBK11081/
30. Pierrot-Deseilligny, Emmanuel and David Burke (2012). The circuitry of the human spinal cord: Spinal and corticospinal mechanisms of movement. Cambridge University Press, pp. 1–606. ISBN : 9781139026727. DOI :10 . 1017 / CBO9781139026727. URL:/core/books/circuitry-of-the-human-spinal-cord/23234E0D24C9A478528747356169FF55.
31. Jankowska, Elzbieta (2008). Spinal interneuronal networks in the cat: Elementary components. DOI : 10.1016/j.brainresrev.2007.06.022. URL : https://pubmed.ncbi.nlm.nih.gov/17884173/.
32. Hallett, M., Shahani, B.T., & Young, R.R. (1977). Analysis of stereotyped voluntary movements at the elbow in patients with Parkinson's disease. Journal of Neurology, Neurosurgery and Psychiatry, 40,1129-1 135.
33. Hannaford, B., Cheron, G., & Stark, L. (1985). Effects of applied vibration on triphasic electromyographic patterns in neurologically ballistic head movements. Experimental Neurology, 88,447-460
34. Shevtsova, N. A., Hamade, K., Chakrabarty, S., Markin, S. N., Prilutsky, B. I., & Rybak, I. A. (2016). Modeling the Organization of Spinal Cord Neural Circuits Controlling Two-Joint Muscles (pp. 121–162). https://doi.org/10.1007/978-1-4939-3267-2_5
35. G.E. Goslow, E.K. Stauffer, W.C. Nemeth, D.G. Stuart, The cat step cycle: response of muscle spindles and tendon organs to passive stretch within the locomotor range, 1973 Brain Res. 60 Ž 1973 . 35–54.
36. Prochazka, A. (1999). Chapter 11 quantifying proprioception. Prog. Brain Res. 123, 133–142.
37.) Rexed, B. (1954). A cytoarchitectonic atlas of the spinal cord in the cat. J. Comp. Neurol. 100, 297–379.
38. Sengul, G., Watson, C., Tanaka, I., and Paxinos, G. (2012). Atlas of the Spinal Cord: Mouse, Rat, Rhesus, Marmoset, and Human. Amsterdam: Elsevier Science.
39. Helen J. Hislop, J. M. (2007). Daniels and Worthingham's Muscle Testing: Techniques of Manual Examination. Elsevier.
40. James, Sebastian S. et al. (2018). "Integrating brain and biomechanical models-A new paradigm for understanding neuro-muscular control". In: Frontiers in Neuroscience 12.FEB, p. 39. ISSN : 1662453X. DOI : 10.3389/fnins.2018.00039. URL:/pmc/articles/PMC5808253/?report=abstract
41. Holash, Robert John and Brian R. MacIntosh (2019). "A stochastic simulation of skeletal muscle calcium transients in a structurally realistic sarcomere model using MCell". In: PLOS Computational Biology 15.3. Ed. by C David Williams, e1006712. ISSN : 1553-7358. DOI : 10.1371/journal.pcbi.1006712. URL : https://dx.plos.org/10.1371/journal.pcbi.1006712.
42. Izhikevich, E. M., & Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical

systems. Proceedings of the National Academy of Sciences of the United States of America, 105(9), 3593–3598. https://doi.org/10.1073/pnas.0712231105
43. Graybiel A. M. (2008). Habits, rituals, and the evaluative brain. Annu. Rev. Neurosci. 31, 359–387. 10.1146/annurev.neuro.29.051605.112851
44. Mallampalli K., Pithapuram, M V., Rangayyan, Y M., Iyengar, R S., Singh, A K., and Raghavan, M (2021). Neuro-musculoskeletal Upper Limb in-silico as virtual patient. BioRxiv, 1. https://doi.org/10.1101/2021.05.16.444298
45. Ting, L. H., & Macpherson, J. M. (2005). A limited set of muscle synergies for force control during a postural task. Journal of Neurophysiology, 93(1), 609–613. https://doi.org/10.1152/jn.00681.2004
46. Flash, T. and Hogan, N. (1985) The coordination of arm movements: an experimentally confirmed mathematical model. J. Neurosci. 5, 1688–1703
47. Uno, Y. et al. (1989) Formation and control of optimal trajectory in human multijoint arm movement. Minimum torque-change model. Biol. Cybern 61, 89–101
48. Kim, T., Hamade, K. C., Todorov, D., Barnett, W., Capps, R. A., Latash, E. M., Markin, S. N., Rybak, I. A., & Molkov, Y. I. (2017). Reward based motor adaptation mediated by basal ganglia. Frontiers in Computational Neuroscience, 11, 19. https://doi.org/10.3389/fncom.2017.00019
49. Leib, R., Russo, M., d'Avella, A., & Nisky, I. (2020). A bang-bang control model predicts the triphasic muscles activity during hand reaching. Journal of Neurophysiology, 124(1), 295–304. https://doi.org/10.1152/jn.00132.2020
50. Jaśkowski, W., Lykkebø, O. R., Toklu, N. E., Trifterer, F., Buk, Z., Koutník, J., & Gomez, F. (2018). Reinforcement Learning to Run... Fast (pp. 155–167). Springer, Cham. https://doi.org/10.1007/978-3-319-94042-7_8
51. Huang, Z., Zhou, S., Zhuang, B., & Zhou, X. (2017). Learning to Run with Actor-Critic Ensemble. ArXiv. http://arxiv.org/abs/1712.08987
52. Kidziński, Ł., Mohanty, S. P., Ong, C., Huang, Z., Zhou, S., Pechenko, A., Stelmaszczyk, A., Jarosik, P., Pavlov, M., Kolesnikov, S., Plis, S., Chen, Z., Zhang, Z., Chen, J., Shi, J., Zheng, Z., Yuan, C., Lin, Z., Michalewski, H., ... Delp, S. (2018). Learning to Run challenge solutions: Adapting reinforcement learning methods for neuromusculoskeletal environments. http://arxiv.org/abs/1804.00361
53. Sutton RS, Barto AG. Introduction to Reinforcement Learning. 1st ed. Cambridge (MA): MIT Press; 1998.
54. M. Botvinick, J. X. Wang, W. Dabney, K. J. Miller, Z. Kurth-Nelson, Deep reinforcement learning and its neuroscientific implications. Neuron 107, 603–616 (2020)
55. Gottlieb, G. L. (1998). Muscle Activation Patterns During Two Types of Voluntary Single-Joint Movement. Journal of Neurophysiology, 80(4), 1860–1867. https://doi.org/10.1152/jn.1998.80.4.1860
56. Morrison, S., & Anson, J. G. (1999). Natural goal-directed movements and the triphasic EMG. Motor Control, 3(4), 346–371. https://doi.org/10.1123/mcj.3.4.346
57. Holzbaur, K.R.S., Murray, W.M., Delp, S.L. A Model of the Upper Extremity for Simulating Musculoskeletal Surgery and Analyzing Neuromuscular Control. Annals of Biomedical Engineering, vol 33, pp 829-40, 2005
58. Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). OpenAI Gym. http://arxiv.org/abs/1606.01540
59. Fuglevand, A. J., Winter, D. A., and Patla, A. E. (1993). Models of recruitment and rate coding organization in motor-unit pools. J. Neurophysiol. 70, 2470–2488. doi: 10.1152/jn.1993.70.6.2470
60. Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R. P., de Berker, A., Ganguli, S., Gillon, C. J., Hafner, D., Kepecs, A., Kriegeskorte, N., Latham, P., Lindsay, G. W., Miller, K. D., Naud, R., Pack, C. C., ... Kording, K. P. (2019). A deep learning framework for neuroscience. In Nature Neuroscience (Vol. 22, Issue 11, pp. 1761–1770). Nature Publishing Group. https://doi.org/10.1038/s41593-019-0520-2
61. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. In arXiv. arXiv. https://arxiv.org/abs/1707.06347v2
62. Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. ACM International Conference Proceeding Series, 382, 1–8. https://doi.org/10.1145/1553374.1553380
63. Angel RW . Electromyography during voluntary movement: The two-burst pattern. Electroencephalography and Clinical Neurophysiology 36: 493-498, 1974

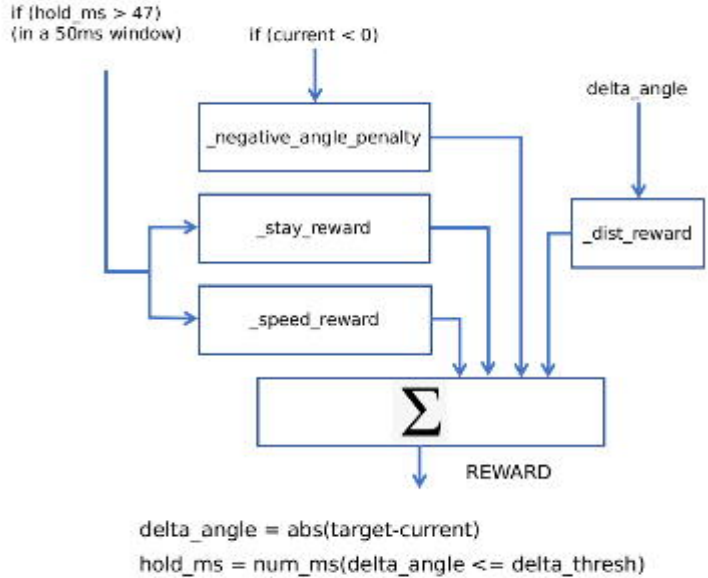
64. Hallett M, and Marsden C . Ballistic flexion movements of the human thumb. The Journal of Physiology 294: 33, 1979.
65. Mustard BE, and Lee RG . Relationship between EMG patterns and kinematic properties for flexion movements at the human wrist. Experimental Brain Research 66: 247-256, 1987.
66. Overduin, S. A., d'Avella, A., Carmena, J. M., & Bizzi, E. (2012). Microstimulation Activates a Handful of Muscle Synergies. Neuron, 76(6), 1071–1077. https://doi.org/10.1016/j.neuron.2012.10.018
67. Rybak, I. A., Dougherty, K. J., & Shevtsova, N. A. (2015). Organization of the mammalian locomotor CPG: Review of computational model and circuit architectures based on genetically identified spinal interneurons. In eNeuro (Vol. 2, Issue 5, pp. 69–84). Society for Neuroscience. https://doi.org/10.1523/ENEURO.0069-15.2015
68. McCrea, D. A., & Rybak, I. A. (2008). Organization of mammalian locomotor rhythm and pattern generation. In Brain Research Reviews (Vol. 57, Issue 1, pp. 134–146). NIH Public Access. https://doi.org/10.1016/j.brainresrev.2007.08.006
69. Kural, Mustafa Aykut., Fabricius, Martin., Christensen, Jakob., Kaplan, Peter W., Beniczky, Sándor. (2020). Triphasic Waves Are Generated by Widespread Bilateral Cortical Networks, Journal of Clinical Neurophysiology: November 20, 2020. doi: 10.1097/WNP.0000000000000770
70. Chen-Bee, C. H., Agoncillo, T., Xiong, Y., & Frostig, R. D. (2007). The triphasic intrinsic signal: Implications for functional imaging. Journal of Neuroscience, 27(17), 4572–4586. https://doi.org/10.1523/JNEUROSCI.0326-07.2007
71. Graziano, M. S. A., Taylor, C. S. R., & Moore, T. (2002). Complex movements evoked by microstimulation of precentral cortex. Neuron, 34(5), 841–851. https://doi.org/10.1016/S0896-6273(02)00698-0
72. Graziano M. The organization of behavioral repertoire in motor cortex. Annual Review of Neuroscience. 29: 105-34. PMID 16776581 DOI: 10.1146/annurev.neuro.29.051605.112924
73. Graziano MS. Ethological Action Maps: A Paradigm Shift for the Motor Cortex. Trends in Cognitive Sciences. PMID 26628112 DOI: 10.1016/j.tics.2015.10.008
74. Grillner S. Control of locomotion in bipeds, tetrapods, and fish. In: Brookhart JM, Mountcastle VB, editors. Handbook of Physiology-The Nervous System II. American Physiological Society; Bethesda, MD: 1981. pp. 1179–1236.


Reinforcement Learning
ACTIONS
OBSERVATIONS
supra-spinal stimulations
Ia, II, Ib
Extensor
Flexor
Neuro D
C6
C5
drg
Alpha Motoneurons
Model Definition
afferent
efferent
(b)
OpenSim
NEUROiD (NEURON+OpenSim)
OpenAI Gym Environment

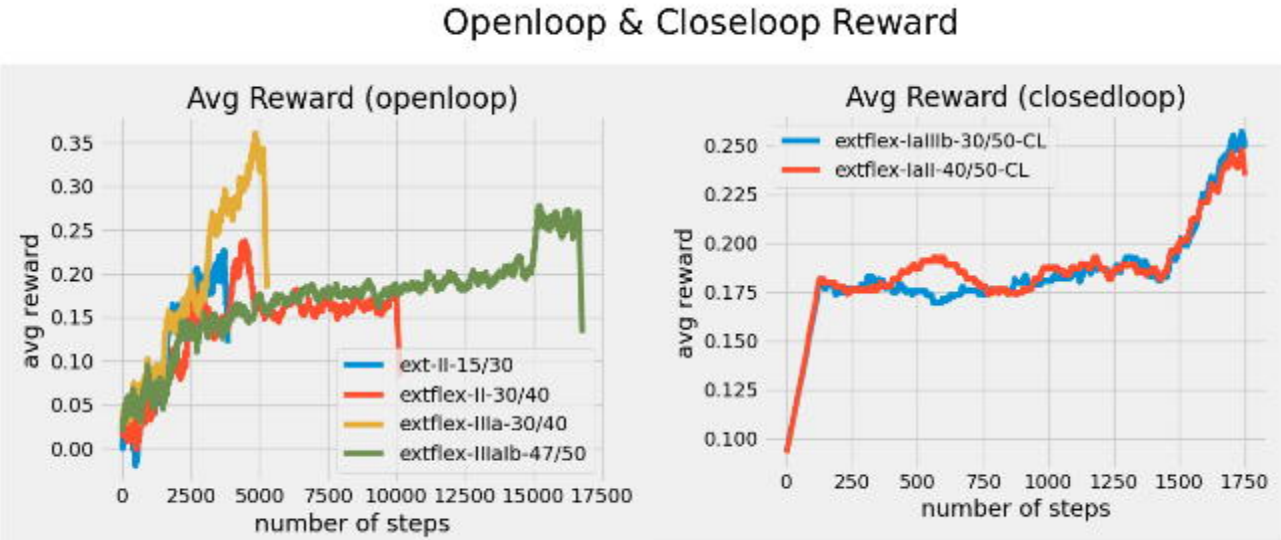

Basal Ganglia
Cerebellum
Descending Systems
Motor Cortex
Brain Stem
Spinal Interneurons
Rhythm Generators
Pattern Formation Network
Spinal Motoneurons and circuits
Musculoskeletal System
Ia, II and Ib
Implemented by RL policy
Not implemented
Implemented in NEURON & OpenSIM
(a)
Descending Tracts
C5
C6
Rhythm Generator
Pattern Formation
IN Network
IaIn
IIIn
αMN
α-MN
IbIn
IbEx
Ia II Ib
Ib II Ia
Biceps
Triceps
Inhibitory
Excitatory
Interneuron (Int & Fire)
α-MN (Realistic)
(b)


[OPENAI GYM]
ENVIRONMENT
[flex_amp, ext_amp]
ACTION
ADV
VALUE
E(Value)
ACTOR
CRITIC
FC
tanh
FC
2
64
64
6
1
64
64
6
PPO
STATE
[flex_{ll,la,lb},ext_{ll,la,lb}]

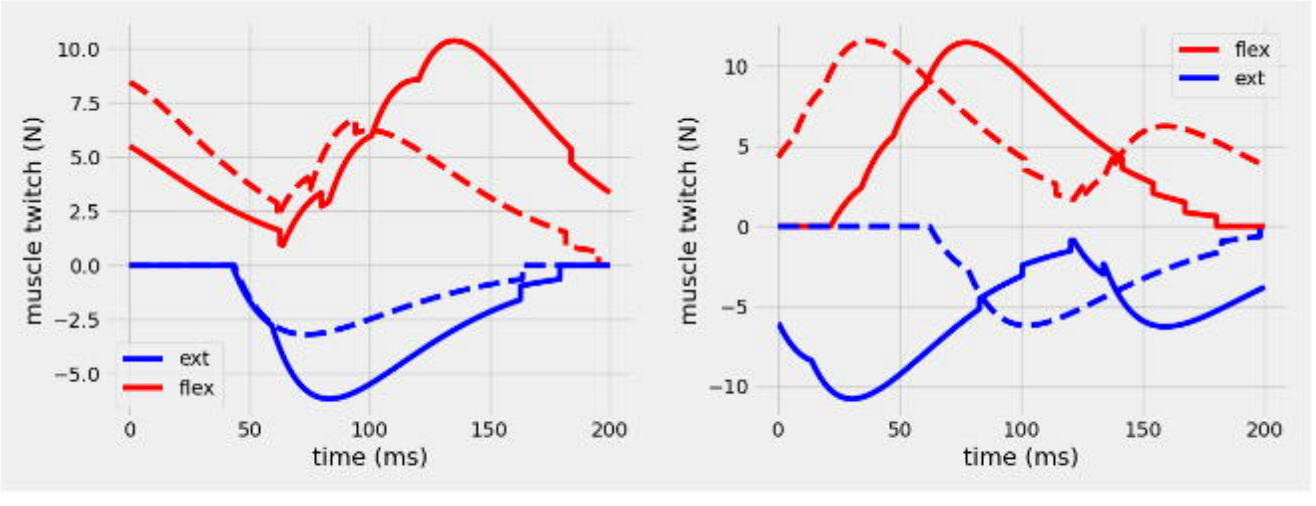
(a)

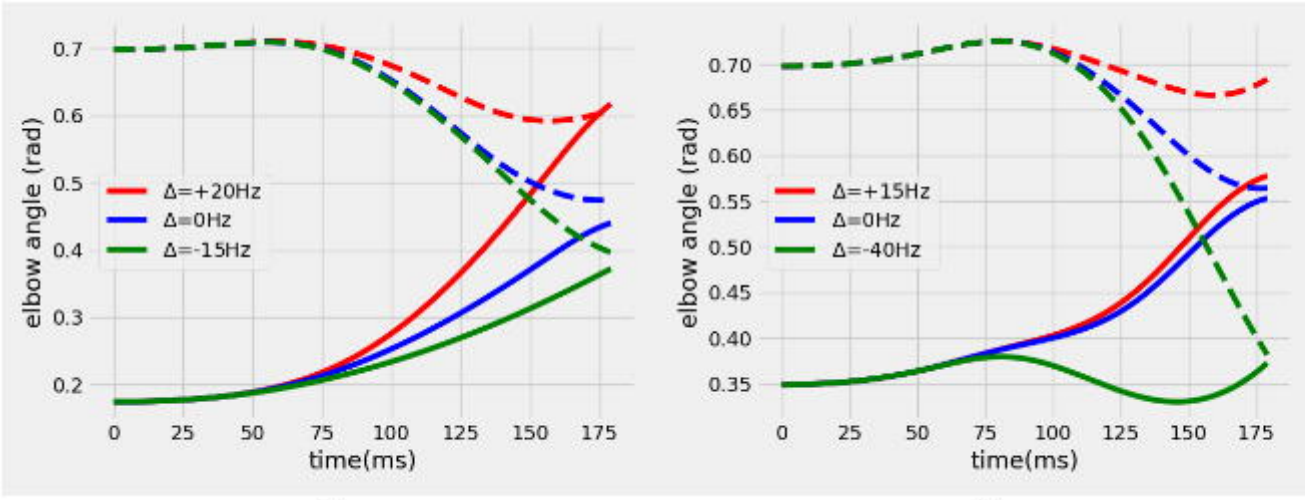

if (hold_ms > 47)
(in a 50ms window)
if (current < 0)
delta_angle
_negative_angle_penalty
_stay_reward
_speed_reward
_dist_reward
Σ
REWARD
delta_angle = abs(target-current)
hold_ms = num_ms(delta_angle <= delta_thresh)

(b)

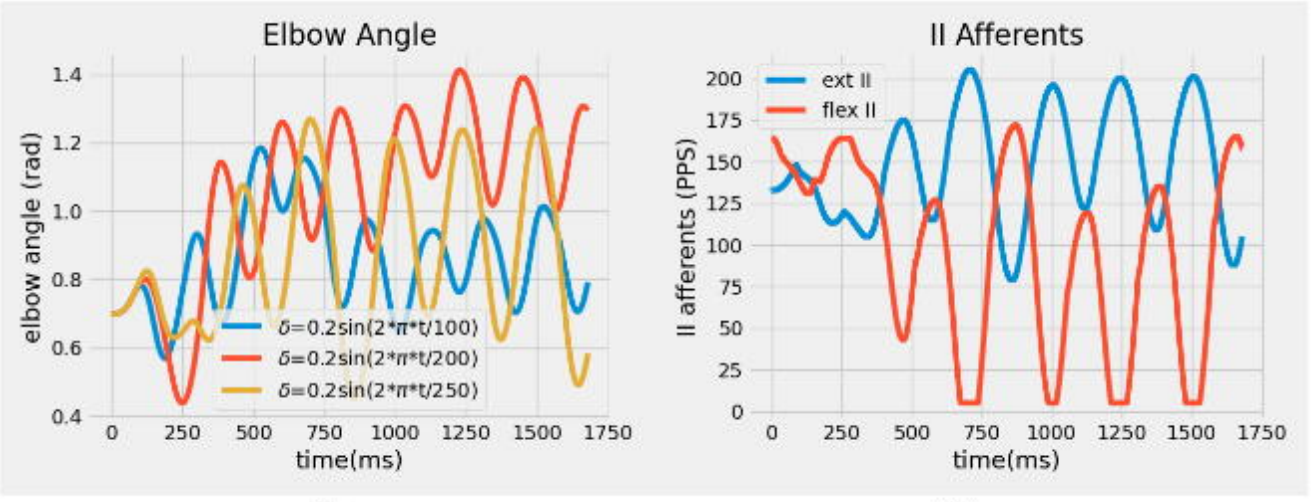

Openloop & Closeloop Reward
Avg Reward (openloop)
avg reward
number of steps
ext-II-15/30
extflex-II-30/40
extflex-IIIa-30/40
extflex-IIIaIb-47/50
(a)
Avg Reward (closedloop)
avg reward
number of steps
extflex-IaIIIb-30/50-CL
extflex-IaII-40/50-CL
(b)

Reward Vs Number of steps
reward
number of steps


10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
muscle twitch (N)
ext
flex
0
50
100
150
200
time (ms)
(a)
10
5
0
−5
−10
muscle twitch (N)
flex
ext
0
50
100
150
200
time (ms)
(b)


elbow angle (rad)
time(ms)
Δ=+20Hz
Δ=0Hz
Δ=-15Hz
(a)
elbow angle (rad)
time(ms)
Δ=+15Hz
Δ=0Hz
Δ=-40Hz
(b)

Musculoskeletal Model
(c)


Elbow Angle
elbow angle (rad)
1.4
1.2
1.0
0.8
0.6
0.4
0
250
500
750
1000
1250
1500
1750
time(ms)
δ=0.2sin(2*π*t/100)
δ=0.2sin(2*π*t/200)
δ=0.2sin(2*π*t/250)
II Afferents
II afferents (PPS)
200
175
150
125
100
75
50
25
0
ext II
flex II
time(ms)

(a) (b)