7th Annual International IEEE EMBS Conference on Neural Engineering
Montpellier, France, 22 - 24 April, 2015

Towards an Integrative Spiking Neuron Model of Motor Control – from Cortex & Basal Ganglia to Muscles & Sensory Feedback

Yufei Wu[1] and A. Aldo Faisal[1,2,3], *Member IEEE*

***Abstract*— We developed an integrative spiking neuron framework to study motor learning and control across multiple levels of biological organisations from synaptic learning rules via neural populations and muscles to an arm's movements. Our framework is designed to simulate reward-based motor learning processes by using identified cellular learning mechanisms (neuromodulation) and enable linking these to findings in human and primate motor learning experiments involving reaching movements. The key learning mechanisms are Actor/Critic reward-based learning and STDP synaptic plasticity rules. We simulate and study learning of planar reaching movements, where motor neuron activities drive Hill-type muscle models which mechanically translate forces into movements via a physics simulator. Our simulated brain is trained and tested in a reaching task with unknown dynamics following a psychophysics protocol. The framework is capable of learning the task and we can directly access the output of neuronal populations (e.g. M1, S1, VTA) as well as EMG-equivalent muscle activations, arm reaching trajectories and sensory feedback before, during and after motor learning. Our ability to simulate and explain motor learning across the levels of neural activity as well as psychophysics experiments will be useful in linking human motor learning experiments to their neuronal correlates. This system can thus provide incisive *in silico* proof-of-principle tests for understanding invasive approaches Brain-Computer-Interfaces and Neuroprosthetics.**

I. INTRODUCTION

The sensorimotor system integrates noisy sensory information and controls noisy skeletal muscle actuators to make body movements – it has to perceive, decide, learn and act in closed-loop [1]. Yet, experimental approaches to understand motor control and learning in humans have to focus on open-loop studies due to methodological limitations in recording motor actions *in vivo* and brain activities (e.g. MEA[2], fMRI[3] but see [4]) Otherwise, closed-loop sensorimotor control can only be accessed by using purely behavioural methods of psychophysics (e.g.[5]). This motivates using simulations through 1. bottom-up modelling and 2. multi-scale stochastic simulation of complex processes [6], [7]. Ultimately perception, action and learning are represented and processed through computational mechanisms embedded in (spiking) neurons. Hence, we need to investigate spiking neuron models, embedding identified reinforcement learning mechanisms in movement task-based learning. Previous work linking neuronal activity and animal behaviour, such as the Morris water maze navigation task [8] and the land-based navigation task with obstacles[9] used two populations of spiking neurons to model hippocampal place cells (representing the state of the agent in space) and the actor in an Actor/Critic framework (representing motor action). In both cases, the actions correspond to abstract directions of locomotion of the animal, not actual muscular controls. The Actor/Critic structure was previously used in control tasks of a one-joint arm, in which the joint angle was directly controlled by spiking neural populations reflecting 'flexor' and 'extensor' activities (but not muscles) [10].

Here, we present an integrative reinforcement learning model linking the cortex to the muscle dynamics. Our system is based on the Actor/Critic structure with spiking neurons for two-joint arm control in planar reaching movements and is thus able to reproduce the benchmark tasks in motor neuroscience. The arm model contains two pairs of flexor and extensor muscles, controlling the shoulder and elbow joints respectively. The mechanics of muscle contractions control tensions and forces on the tendons which in turn move the arm. Our model incorporates both kinematics and dynamics. It thereby supports an effective simulation of motor learning and control in a closed-loop muscle-based system driven by spiking neurons.

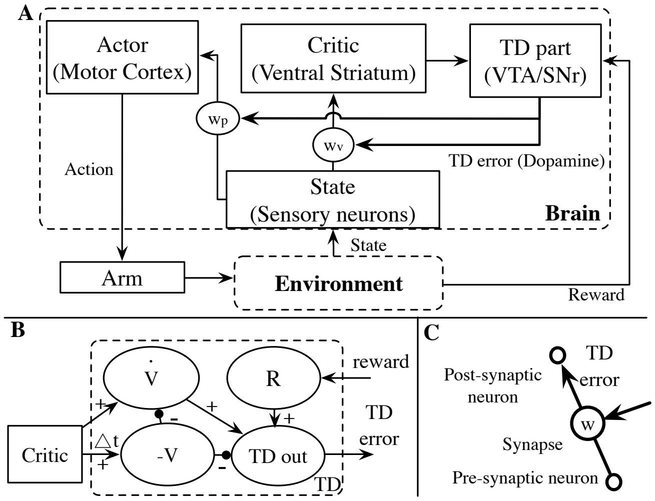

Fig. 1. Integrated spiking neuron model linking the cortex to muscles. (**A**) Overview of interacting neural populations, the arm and the environment state. (**B**) Structure of the temporal difference component. The neural Actor/Critic reinforcement learner implements basal ganglia learning mechanisms. V: state value, R: reward, TD: temporal difference. (**C**) Dopamine-based plasticity. LTP-flavoured learning modulated by dopamine signals from the basal ganglia implements the control policy by learning synaptic weights of projections from sensory state (S1) to critic (ventral striatum) and mapping from S1 to M1 (actor).

Brain & Behaviour Lab - [1]Department of Computing, [2]Department of Bioengineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK & [3]MRC Clinical Sciences Centre, Hammersmith Hospital Campus, W12 0NN London, UK. `a.faisal at imperial.ac.uk`.

978-1-4673-6389-1/15/$31.00 ©2015 IEEE

II. METHODS

a) Sensorimotor system simulation: Simulations involve the physical environment, the biomechanics of the arm, and various sensory, motor and interernal neuronal populations of the brain (Fig. 1 (A)). The latter consists of specific brain areas and the functional relationship between them are specified by fixed or adaptive synaptic connection patterns. Conceptually all learning is driven as reward-based learning using an Actor/Critic model. Sizes of neural populations were set manually to satisfy the accuracy requirements of the biomechanical system and the reaching task given and were as follows: 1. state population("sensory neurons"), Poisson model, 7920 neurons; 2. actor population ("M1"), Leaky-Integrate-Fire model, 256 neurons; 3. critic population ("ventral striatum"), Poisson model, 200 neurons, TD populatioon ("VTA/SNr"), Poisson mode, 800 neurons. Policy encoded as synaptic weight matrix W_p of size 7920 × 256. Value estimator encoded as synpatic weigh matrix W_v of size 7920 × 200.

State representation ("S1"): The neural population of the S1 reflect information from the environment and represent the current state of the arm. The state is defined as $s = (x, y, \dot{x}, \dot{y})$, i.e. the position and the velocity of the wrist in a 2D plane, thus a continuous 4D space. Sensory neurons encode the state using a simple population code[12]: Each state neuron stands for one point in the 4D state space and its fire rate is determined by its distance to the current state, defined as $r_i(s) = r_{max} \exp(-\frac{(x-x_i)^2}{2\sigma_1^2} - \frac{(y-y_i)^2}{2\sigma_2^2} - \frac{(\dot{x}-\dot{x}_i)^2}{2\sigma_3^2} - \frac{(\dot{y}-\dot{y}_i)^2}{2\sigma_4^2})$. $(x_i, y_i, \dot{x}_i, \dot{y}_i)$ is the corresponding point of state neuron i in the state space and $s = (x, y, \dot{x}, \dot{y})$ is the current state of the arm. r_{max} is the maximum firing rate.

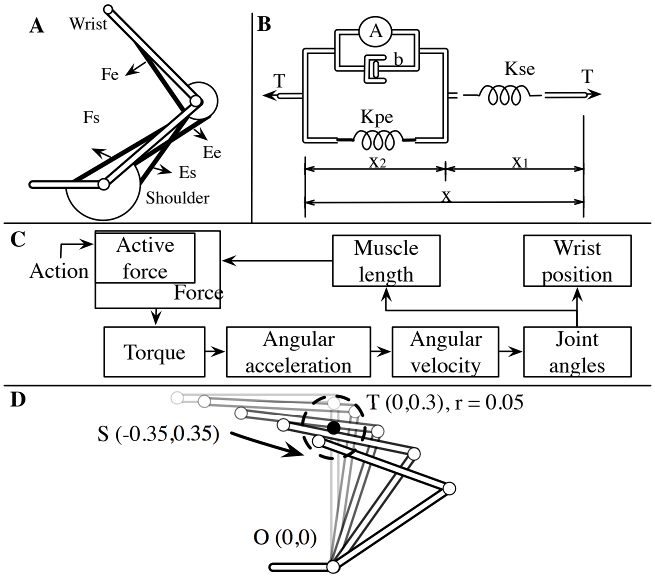

Fig. 2. Muscle-based arm simulation and target-seeking task description. (**A**) Muscle connections. The arm contained two joints which enabled it to move in a two dimensional space. The lengths of the upper and lower arm were both 0.35 m. Fe, Fs: flexor muscles; Es, Ee: extensor muscles. (**B**) Hill-type muscle model. 'A' represents the active force. $K_{SE} = 136g/cm$, $K_{PE} = 75g.cm$ and $b = 50g \cdot s/cm$ (values from [11]) are coefficients of the damper and the springs. T: muscle tension, x: muscle length. (**C**) Arm simulation process containing kinematics and dynamics. (**D**) Description of the target-seeking task. The arm structure was placed in a Cartesian coordinate system with the origin O(0,0) locating on the shoulder (1 unit = 1 m). The starting point was S (-0.35,0.35), whose corresponding joint angle was $(\frac{\pi}{2}, \frac{\pi}{2})$. The target area was a circle around the target T(0,0.3) with target area radius $r = 0.05$.

Actor ("M1") operates as an action generator, encoding control signals for muscles - we ignore the spinal chords role in motor control for the sake of simplicity in this first model of ours. Actions are defined as active forces to drive two pairs of extensor and flexor muscles and are generated based on the population coding model of the primary M1 [13]. Each actor neuron represents a unit vector indicating its preferred direction and preferred directions are evenly arranged from 0 to 2π. There are internal connections between the actor population units which determine a competition process. Each actor neuron has excitatory synapses to neurons with similar preferred directions and inhibitory synapses to all others. The population vector is calculated based on the firing activities of the entire population as $\boldsymbol{P} = [P_x, P_y]^T = \sum_{j=1}^{N_A} r_j \boldsymbol{D}_j$, where r_j and $\boldsymbol{D}_j$ are the firing rate and the direction vector respectively of actor neuron j and N_A is the size of the actor neural population. The action (two pairs of muscle forces) is designated as $E_s = a_1 \max(P_x, 0)$; $F_s = a_1 \max(-P_x, 0)$,$E_e = a_2 \max(P_y, 0)$; $F_e = a_2 \max(-P_y, 0)$, where E_s, F_s, E_e and F_e are corresponding active forces of muscles in Fig. 2 (A) and a_1, a_2 are coefficients. The motor action is updated in 100 ms intervals using the firing acticity of the previous interval.

Critic & TD ("Ventral striatum and VTA/SN")): The ventral striatum represents the value of the current state, while the VTA/SNr computes the reward prediction error, in form of the neuromodulator dopamine. The reward prediction error (or TD error) is calculated following the temporal difference learning in continuous time[14]: $\delta(t) \equiv r(t) - \frac{1}{\tau}V(t) + \dot{V}(t)$, where r represents rewards given by the environment, τ is the time constant for reward discount and V, $\dot{V}$ are the state value and its time derivative. To represent continuous TD learning using neural activities, the TD computations are represented by four neuronal populations (see Fig. 1 (B)), each of which encodes a variable in the TD error equation.

Dopamine-based learning rule: The policy is encoded in the synaptic connection matrix W_p between the S1 (state) and the M1 (actor). The synaptic connection matrix W_v between the S1 (state) and the ventral striatum (critic) represents the reward value-estimator. We use here a dopaminergic long-term potentiation (LTP) to model the synaptic connection dynamics. The individual synaptic weight change (see in Fig 1 (C)) according to the dopamine signal (as defined by the TD error term). The synaptic update are defined as: $\Delta w = \eta \cdot e_{TD} \cdot g(F_{pre}, F_{post})$, where the weight variation of a synapse is determined by the learning rate η, TD error e_{TD} and the LTP parameters $g(F_{pre}, F_{post})$.

b) Biomechanical simulation of the arm: We model a two-joint arm to model planar reaching movements (Fig. 2 (A)). The arm is controlled by two pairs of flexor and extensor muscles connected to the shoulder and the elbow joint (we ignore here for simplicity biarticulate muscles).

Each muscle is simulated using a basic Hill-type muscle model[11], in which the muscle force is the combination of a passive force and an active force (Fig. 2 (B)). The passive muscle force is generated by the springs and the damper in the muscle model whereas the active force follows the control signal given by the actor neurons ("M1") in the sensorimotor system.

Simulation of our muscle-driven arm involves both arm kinematics and dynamics (Fig. 2 (C)). Control signals (actions) driven by the motor neurons, determine the contraction forces of the 4 muscles, which in turn influence the rotation torques of the arm joints. The resulting angular velocities of the joints are directly reflected in movements of the arm. Arm movements change the arm posture, and consequently the 1. muscle lengths (passive forces) and 2. moments of inertia. Muscles lengths affect passive forces in the Hill-type muscle model. Moments of inertia affect how muscles generated accelerations of the arm based on Newton's law of motion. Thus, arm is from a control theoretical perspective a non-linear plant. Arm and muscles together make the system a highly non-linear plant with hysteresis. Ab initio our simulated brain does not know how to control such a system, its kinematics and dynamics have thus to be learned by our framework. Mechanics, M1 drive and force parameters are taken from the literature and adjusted so as to produce realistic human arm movement velocities and forces.

c) Motor Learning: The integrative spiking neuron system is given a task derived from psychophysics learn to reach with the arm from an starting point to a target. All synpatic weights have their initial values set to (illustrated in Fig. 2 (D)) 0.5 for the state-critic synapses and uniform random values between 0.4 to 0.6 for the state-actor synapses. All other synaptic weights are constant. Rewards are delivered according to the square distance between the wrist and the target with an additional large reward once the target area is reached, this is consistent with psychophysics level modelling of reaching tasks. Trials in which the wrist reaches the target area are considered successful. We define the cost of each successful trial: $COST = T_{used}/D_{min}$, where T_{used} is the consumed time (second) and D_{min} is the minimum distance (meter) to reach the target. A trial in which the task is not completed within a spatial or temporal limit is considered a missed trial and restarted. Our neuronal framework has to learn to move so as to maximise reward and minimise costs.

III. RESULTS

Our integrative spiking neuron system can learn to move: performance in the reaching task is improved after the learning process. The median of task cost (over 50 movements) decreases from 18.2 to 5.5 after only 50 reachign trials.Fig. 3 shows a 1 second long snapshot of the neural activities in the populations as well as the arm movement trajectory. We chose the initial first second from trial start and display snapshot for both, before learning (initial setting) and after learning (50 trials). Before the learning process, activities of actor neurons (M1) are determined by our arbitrary weight initialisations and unstructured firing patterns of state (sensory) neurons. Their firing pattern consists of firing activities of 'winners' in the competition and each pattern has a similar chance to appear. As shown in Fig 3 (B), the firing pattern of actor neurons is very unstable and there can be large differences between actions, even between very similar arm states. In contrast, actor neuron activity after learning in Fig. 3 (C) shows a clear and stable structure, with gradual changes between similar states. This reflects the control policy that was learned, which control the arm to move efficiently towards the target area. The state value is represented by the mean firing rate of critic neurons (ventral striatum): before learning, these fire at high rates regardless of the arm's state. After learning, the mean firing rate of critic neurons increases as the arm approaches the target area, thus having learned a representation of the experimenter controlled task reward.

IV. CONCLUSION

We built a framework for simulating motor learning processes of the brain using a spiking neuron system that can learn to control highly-nonlinear reaching movement tasks. We demonstrate the ability of our learning system to learn an appropriate policy for the task. Our model implements reward-based learning, sensory and motor representations as spiking neuronal populations with neuromodulated synapses.The model includes mapping from cortex to muscles and back (sensory feedback). It thereby provides a first step towards better understanding of cortical activity and neural interfacing and can inform the design and development of Brain Computer Interfaces. In contrast to previous work, our approach involves 1. bottom-up modeling - aggregating from first-principles published data to model function at a higher level of organisation (e.g. muscle contractions explain movement dynamics) and 2. multi-scale modeling - operating at multiple levels of biological organisation, where functional elements use more microscopic rules (e.g. here abstract Actor-Critic learning is mapped to synaptic plasticity rules). These two approaches have been successful in explaining diverse high-level properties of the brain, e.g. [15], [16], or disease conditions, e.g. [17], [18]. While our approach is on the one hand simplistic in capturing only selected elements of the sensorimotor control loop, it captures many elements of the perception-action loop including muscle dynamics and arm mechanics typically neglected in computational neuroscience. We will incorporate further aspects to model the dynamics of spinal chord [19], cerebellum [20] and more realistic synaptic plasticity rules [21]. Our aim is to use this system to explore what systems and mechanics at the cellular and system level are required to reproduce psychophysics level experiments in motor learning. To deal with practical motor learning tasks, this model needs to be improved with additional learning mechanisms (eg. error-based learning). Vice versa, given high-level strategies and representations for a given learning task, we can test if our neuronal representations match electrophysiological or imaging data. The finding that our simplistic, holistic system of spiking neuron populations can learn to control the non-linear dynamics of 4 muscles and mechanics of a 2 joint arm to learn a reaching task in 50 trials, shows that in principle, our approach bears the propmise to understand motor control and learning in an integrative spiking neuron system.

References

[1] A. A. Faisal, L. P. Selen, and D. M. Wolpert, "Noise in the nervous system," *Nature Rev Neurosci*, vol. 9, no. 4, pp. 292–303, 2008.

[2] Q. Fu, J. Suarez, and T. Ebner, "Neuronal specification of direction and distance during reaching movements in the superior precentral premotor area and primary motor cortex of monkeys," *J Neurophysiol*, vol. 70, pp. 2097–2097, 1993.

[3] E. Hecht, D. Gutman, N. Khreisheh, S. Taylor, J. Kilner, A. Faisal, B. Bradley, T. Chaminade, and D. Stout, "Acquisition of paleolithic toolmaking abilities involves structural remodeling to inferior frontoparietal regions," *Brain, Structure & Function*, pp. 1–17, 2014.

[4] A. Sylaidi, P. Lourenco, S. Nageshwaran, C.-H. Lin, M. Rodriguez, R. Festenstein, and A. A. Faisal, "f2MOVE: fMRI-compatible haptic object manipulation system for closed-loop motor control studies," *Proc.7th Intl IEEE/EMBS Conf on Neural Engineering (NER)*, 2015.

[5] A. A. Faisal and D. M. Wolpert, "Near optimal combination of sensory and motor uncertainty in time during a naturalistic perception-action task," *J Neurophysiol*, vol. 101, no. 4, pp. 1901–1912, 2009.

[6] A. A. Faisal, "Stochastic simulation of neurons, axons and action potentials," in *Stochastic Methods in Neuroscience*, pp. 297–343, Oxford University Press, 2010.

[7] A. A. Faisal, "Noise in neurons and other constraints," in *Computational Systems Neurobiology*, pp. 227–257, Springer Netherlands, 2012.

[8] E. Vasilaki, N. Frémaux, R. Urbanczik, W. Senn, and W. Gerstner, "Spike-based reinforcement learning in continuous state and action space: when policy gradient methods fail," *PLoS Comp Biol*, vol. 5, no. 12, p. e1000586, 2009.

[9] N. Frémaux, H. Sprekeler, and W. Gerstner, "Reinforcement learning using a continuous time actor-critic framework with spiking neurons," *PLoS Camp Biol*, vol. 9, no. 4, p. e1003024, 2013.

[10] 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*, vol. 7, no. 10, p. e47251, 2012.

[11] R. Shadmehr, *The computational neurobiology of reaching and pointing: a foundation for motor learning*. MIT press, 2005.

[12] P. Dayan and L. F. Abbott, *Theoretical neuroscience*. Cambridge, MA: MIT Press, 2001.

[13] A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner, "Neuronal population coding of movement direction," *Science*, vol. 233, no. 4771, pp. 1416–1419, 1986.

[14] K. Doya, "Reinforcement learning in continuous time and space," *Neural computation*, vol. 12, no. 1, pp. 219–245, 2000.

[15] B. Sengupta, A. A. Faisal, S. B. Laughlin, and J. E. Niven, "The effect of cell size and channel density on neuronal information encoding and energy efficiency," *J Cereb Blood Flow & Metab*, vol. 33, no. 9, pp. 1465–1473, 2013.

[16] A. Neishabouri and A. A. Faisal, "Axonal noise as a source of synaptic variability," *PLoS Comp Bio.*, vol. 10, no. 5, p. e1003615, 2014.

[17] A. Neishabouri and A. A. Faisal, "Saltatory conduction in unmyelinated axons: clustering of Na+ channels on lipid rafts enables micro-saltatory conduction in C-fibers," *Front Neuroanat*, vol. 8, p. 109, 2014.

[18] A. A. Faisal, "Studying channelopathies at the functional level using a system identification approach," *CompLife*, vol. 940, pp. 113–126, 2007.

[19] G. A. Tsianos, J. Goodner, and G. E. Loeb, "Useful properties of spinal circuits for learning and performing planar reaches," *J Neural Eng*, vol. 11, no. 5, p. 056006, 2014.

[20] "Neuromodulatory adaptive combination of correlation-based learning in cerebellum and reward-based learning in basal ganglia for goal-directed behavior control," *Frontiers Neural Circ*, vol. 8, no. 00126, 2014.

[21] R. Legenstein, D. Pecevski, and W. Maass, "A learning theory for reward-modulated spike-timing-dependent plasticity with application to biofeedback," *PLoS Comp Biol*, vol. 4, no. 10, p. e1000180, 2008.

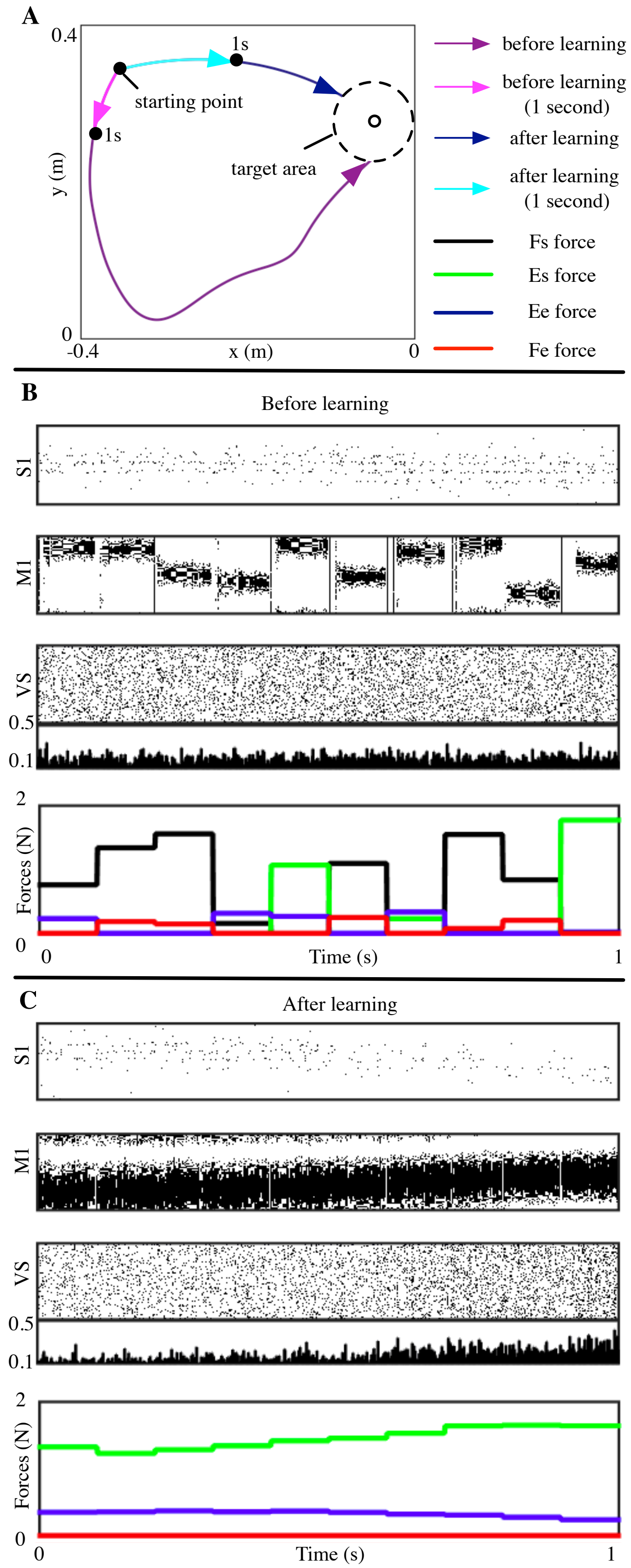

Fig. 3. (**A**) Reaching trajectories before and after motor learning. Light purple arrow: trajectory in the first second before learning, corresponding neural activities are shown in (B). Light blue arrow: trajectory in the first second after learning, corresponding neural activities are shown in (C). (**B-C**) Neural activities and muscle forces in the first second. S1: part of state neurons (S1). M1: actor neurons (M1). VS: critic neurons (ventral striatum) with mean firing rate histogram. Fe, Fs: flexor muscles; Es, Ee: extensor muscles. (see Fig. 2 (A))