2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Neural Encoding of Reaches in a Linear Cortical Model

Patrick Greene[1], Marc H. Schieber[2], and Sridevi V. Sarma[1], *IEEE Senior Member*

Abstract— **To effectively control the arm, motor cortical neurons must produce complex patterns of activation that vary with the position and orientation of the arm and reach direction. In order to better understand how such a finely tuned dynamical system could arise and what its basic organizing principles are, we develop a model of the motor cortex as a linear dynamical system with feedback coupled to a two-joint model of the macaque arm. By optimizing the connections between neural populations with respect to an objective function that penalizes error between hand and target, as well as neural and muscular energy use, we show that certain properties of the motor cortex, such as muscle synergies, can naturally be obtained. We also demonstrate that the optimization process produces a stable neural system in which targets in the physical space are mapped to attracting fixed points in the neural state space. Finally, we show that this optimization process produces neural units with complex spatial and temporal activation patterns.**

I. INTRODUCTION

There is significant debate over the encoding and organizational principles that underlie cortical control of movement in healthy brains [1]. Traditionally, strong experimentally observed correlations have suggested that neural activity encodes directly observable variables like hand or arm position, velocity, or acceleration within a fixed reference frame [2]. However, recent work suggests that the preferred direction of M1 neurons tends to reflect activity in muscle space rather than the extrinsic hand space [3]. This view is supported by spike-triggered averaging of myographic activity, which shows that many M1 neurons have direct connections to motoneurons, both excitatory and inhibitory [4].

Most neurons that have such connections activate multiple synergistic muscles [4]. Each muscle is represented across many neurons with its own set of synergies, so that, even within the same set of muscles, different populations of neurons will be active depending on the initial posture and temporal order of muscle activation [5]. These synergies tend to be combinations of muscles that span multiple joints [6]. How synergies are selected by neurons and how their disruption in injured brains affects recovery are current topics of research [7].

Beyond synergies, the larger picture of how the motor cortex computes temporal patterns of muscle activation that produce efficient trajectories for reach targets regardless of their spatial position or movement direction is still unknown. One promising approach views the activity of the motor cortex through the lens of dynamical systems theory, and posits that computation is done as the neural dynamics moves through the neural state space [8]. The way it moves through this space is determined by the fixed points in the space and their properties - whether they attract, repel, or otherwise interact in different ways to produce the overall ”flow field” that moves the dynamics along [8]. The locations and properties of these fixed points is in turn determined by the highly tuned connections between neurons, which raises the question of how the brain is able to find connections that produce the desired dynamics.

One possibility is that the required connections can be found through an optimization process. In nature, evolutionary pressures over millions of years have shaped the neural architecture of the primate motor cortex in a way that optimizes survival. Over much shorter time scales, synaptic plasticity and learning in young children allows them to develop the fine grained motor skills needed for accurate reaching.

In this paper we develop a simple model of the motor cortex and the arm. In contrast to previous models such as [9]–[11], we use a two-joint, physics-based model of the arm, and the parameters of our cortical model are optimized for general control of the arm rather than being fit to experimental data for a particular task. We show that certain properties of neurons, muscles, and the motor cortex as a whole arise naturally as a result of optimizing reach efficiency and accuracy over a broad set of reaches. The optimized neuron to muscle connections exhibit basic features of muscle synergies, and targets in the physical space are mapped to attracting fixed points in the neural space in a simple manner. Finally, we show that this optimization process produces neural units with activation patterns that can be either mono-, bi-, or triphasic in time, and that these activation patterns depend on both the direction of reach as well as the region in space in which the reach is occurring.

II. METHODS

A. Arm model

We constructed a two-dimensional, physics-based arm model with two joints and four muscles. There are two arm segments, whose masses and lengths were set to typical values of the Rhesus macaque (*Macaca mulatta*), taken from [12]. The joints consist of a shoulder and elbow joint, each of which has its angular motion constrained to realistic ranges.

Joint torques in the arm model are produced by a set of four uniarticular muscles - two shoulder muscles and two elbow muscles, one flexor and one extensor in each case.

[1] Institute for Computational Medicine, Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA

[2] Department of Neurology, University of Rochester, Rochester, NY, USA
Correspondence: pagreene@jhu.edu

This work was supported by NIH/NINDS grant R01NS110423

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

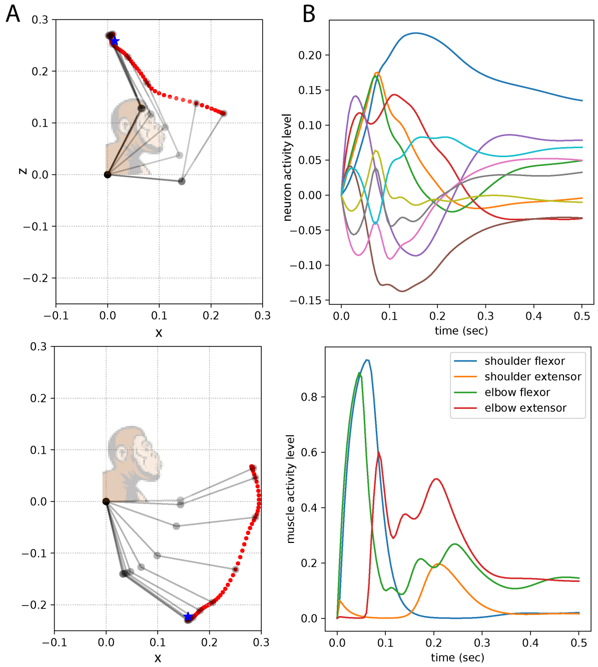

Fig. 1. A) Example reaches of the arm and motor cortex model in two different directions. The target is indicated by a blue star and the position of the arm is shown every 50 ms. The red dots are plotted every 5 ms and indicate the path of the hand. B) Top: Activity levels for each of the 10 neural units for the top reach shown in A). Bottom: Corresponding muscle activity for the top reach shown in A)

Muscle moment arms were taken from [13]. For simplicity and to improve simulation speed, muscles are assumed to have linear dynamics of the form $\dot{m} = \frac{1}{\tau_m}(-m + I_n)$ where m is the muscle activation, $\tau_m = 0.02$ sec is the activation (and deactivation) time constant, as modeled in [13], [14], and I_n is the neural input function which will be made explicit below. For a given muscle activation level, the torque about the spanned joint is calculated by multiplying the muscle activation by the maximum muscle force and the moment arm.

The kinematics and equations of motion for the arm were derived using standard techniques (see i.e. [13], [14]). The equations of motion will be denoted $\dot{r} = R(r,m)$, where $r = [q_1\ q_2\ \dot{q}_1\ \dot{q}_2]$ is the vector of joint angles and angular velocities.

B. Neural model

The arm model is controlled using a linear dynamical system with feedback. The set of neural units has some firing rate at time t denoted by the vector $x(t)$, and all units receive input from other neural units with weights given by the connectivity matrix A. They also receive a bias input B_0, target input $B_1 u_1$, where u_1 is the two-dimensional target vector and B_1 is a matrix mapping the target input to the neural state space, hand position feedback of the form $C_1 v_1(r)$, where $v_1(r)$ is the position of the hand at any given time, and muscle feedback of the form $C_2 m$.

The neural model is coupled to the muscle dynamics via the neural input function $I_n = \sigma(D_0 + D_1 x)$ where D_1 is the neural unit to muscle connectivity matrix, D_0 is a bias term, and $\sigma(x) = 1/(1+e^{-x})$ maps the input to between 0 and 1. In summary, the dynamics of our coupled motor cortex and arm model are given by:

$$\begin{aligned}\dot{x} &= Ax + B_0 + B_1 u_1 + C_1 v_1(r) + C_2 m\\ \dot{m} &= \frac{1}{\tau_m}(-m + \sigma(D_0 + D_1 x))\\ \dot{r} &= R(r,m)\end{aligned} \tag{1}$$

Because each unit in our model has a smooth output that represents a mean-subtracted firing rate and can have both positive and negative connections to other units, we consider the units as representing populations of neurons with similar sets of inputs and activation patterns. We therefore use the terms ”units” or ”neural units” to refer to them.

C. Optimization

Given a training set of S samples consisting of starting points and targets spread uniformly over the reachable space and a time window of length T in which to make each of the reaches, we optimized the matrices A, B_0, B_1, C_1, C_2, D_0, D_1 simultaneously in our model with respect to the following objective function:

$$\sum_{i=0}^{S} \int_0^T \|pos(r) - u_{1,i}\|^2 + \beta_2 \|x\|^2 + \beta_3 \|m\|^2 \, dt \tag{2}$$

where $pos(r)$ is the time-varying position of the hand and $u_{1,i}$ is the target for sample i. The first term in the integral measures the accuracy of the reach, the second term $\|x\|^2$ measures the amount of neural activity required during the reach, and the third term $\|m\|^2$ measures the amount of muscular energy consumed. The factors β_2 and β_3 determine the relative importance of minimizing each of these three terms. Note that while the target $u_{1,i}$ is constant during a given reach, the hand and muscle feedback are continuously changing as the arm moves, resulting in time-varying motor commands due to this feedback and the connections between neural units. The optimization was accomplished using Pontryagin's adjoint sensitivity method [15].

III. RESULTS

We ran the above optimization for models ranging from 4 to 40 units, $T = 1$ sec, $S = 1000$ samples, and a wide range of penalty parameters. We found that as long as the number of units is approximately 10 or more and β_2 and β_3 are kept below about $5 \times 10^{-2}/N$ and 5×10^{-3}, respectively (where N is the number of units), the results are consistent across different numbers of neurons or values of the penalty parameters. In this paper we will show neural activity and connection patterns for a 10 neuron model which was optimized using $\beta_2 = 10^{-3}$ and $\beta_3 = 2.5 \times 10^{-3}$. Figure 1 shows two example reaches using the optimized model, along with associated neural population and muscle activity.

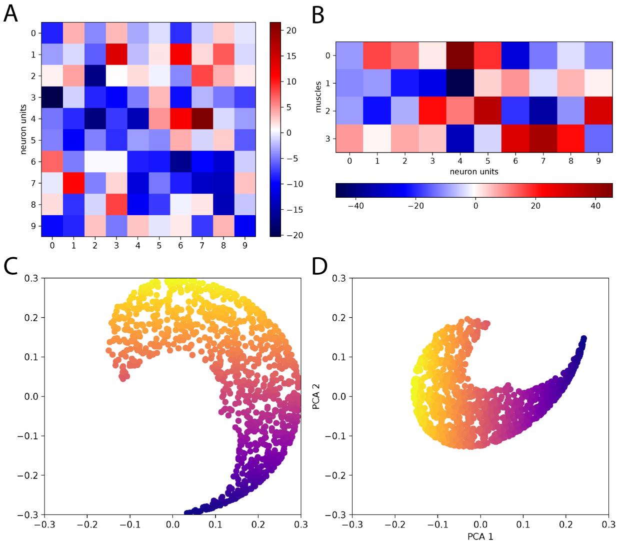

Fig. 2. A) Matrix of connections between neural units. Entry i,j indicates the connection strength between unit i and unit j. B) Matrix of connections between neural units and muscles. Column i indicates the connection strength between unit i and each of the four muscles, with 0 = shoulder flexor, 1 = shoulder extensor, 2 = elbow flexor, 3 = elbow extensor. C) Set of reach targets for which the steady state neural activity was calculated. Coloration is arbitrary and for reference to D). D) Projection of steady state neural activity on to first two PCA components for the set of targets shown in C), with each point in PCA space colored according to the corresponding target presented in C), showing that there is no mixing.

A. *Neural unit connections*

The connections across neural units is shown by the connectivity matrix in Figure 2 A). Entry i, j indicates the amount that unit i weights input from unit j. Eight of the 10 eigenvalues of this connectivity matrix are complex conjugate pairs, which give rise to oscillatory modes in the neural activity at frequencies of 4.04 Hz, 2.08 Hz, 1.74 Hz, and 0.24 Hz. The real parts of all the eigenvalues are negative after optimization, a property whose consequences we will discuss further in section C). Reaches typically take between 0.2 and 0.5 seconds, and consist of an initial application of force which produces an acceleration of the hand, followed by a deceleration phase starting from roughly midway through the reach, with perhaps some additional force at the end required for correction or to hold the arm position. Thus, the oscillatory frequencies we observe in our connection matrix are in line with the required biphasic or triphasic modulation of the muscle within a few tenths of a second needed to produce efficient trajectories.

B. *Neural unit to muscle connections*

The connections between neural units and muscles after optimization are shown in Figure 2 B). Column i indicates the connection strength between unit i and each of the four muscles. A negative connection means that positive activity by the neural unit drives the muscle toward deactivation, while negative activity activates it (and vice versa for positive connections). We note that the optimized neuron to muscle connections exhibit basic features of muscle synergies in that the connections tend to span multiple joints - most units have positive connections to one of the shoulder muscles and one of the elbow muscles and negative connections to the opposing muscles, shown by two red squares and two blue squares in each column. Antagonistic muscles are almost never coactivated by the same neuron; either both muscles are deactivated or one muscle is activated while the other is simultaneously deactivated, resulting in efficient movement without unnecessary muscle activation and wasted energy.

C. *Fixed points in the neural state space*

To investigate how targets are mapped into the space of neural activity, we note that since the real parts of all the eigenvalues of the connectivity matrix are negative, the neural dynamics are strictly stable and every input target produces a single attracting fixed point in the neural state space. This fixed point does not depend on the initial conditions of the system such as the starting position of the hand. In Figure 2 C), we plot a set of 1000 target positions within the reachable space of the arm (as usual the shoulder, not shown, is at (0,0)). For each of these target positions we calculate the corresponding fixed point in the neural state space. Since these points are 10 dimensional, we then apply principle components analysis to project these fixed points into a 2 dimensional subspace that we can visualize. The result is shown in Figure 2 D), with the color of each fixed point corresponding to the color of the target from which it was calculated in C). As we can see, the targets are mapped into the fixed points of the neural state space with almost no distortion or mixing. Mathematically, this is thanks to the linearity of the motor cortex model and the fact that the sigmoid in the muscle activation equation provides only a mild non-linearity. As we described above, each of these fixed points is an attractor in the neural state space, and thus at a high level we can describe the operation of our motor cortex model as mapping targets to points in the neural state space and using attractor dynamics to move the arm toward the target.

D. *Neural activation fields*

In order to understand in detail what aspects of movement each neural unit is coding for, we consider 3 different factors in any given reach: 1) the region of space in which the reach occurs, 2) the direction of the reach, and 3) the temporal course of activation. We subdivide the reachable area into 4 regions and place a target at the center of each region. For each target, we record neural activity in our model as it reaches for the target, starting from a ring of 32 initial hand positions located 7 cm from the target (or 5 cm in the case of the lower-most region due to the shape of the reachable space). We then calculate the fraction of the total neural activity contributed by each neural unit at each point in time by taking the squared valued of the unit's activity and dividing by the sum of squares of the activity of all the units in the model. The result is shown in Figure 3 A). In each plot, the physical hand space is shown as in Figure 1, with the shoulder (not shown) at (0,0).

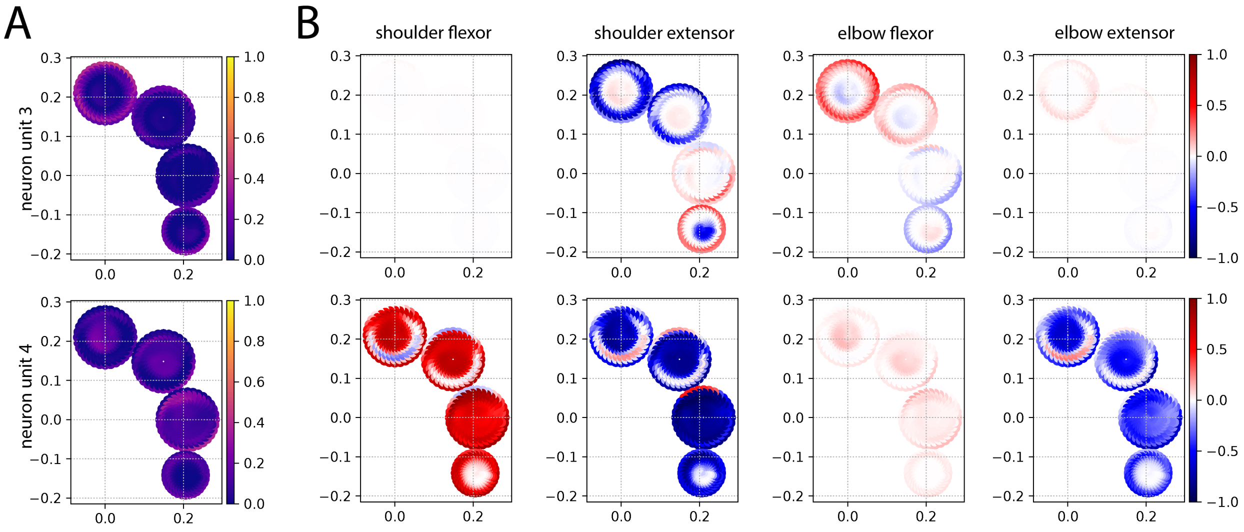

Fig. 3. Relative contribution of neural units to A) the total neural activity and B) the total muscle input for each muscle. In each plot, the physical hand space is shown as in Figure 1. There are 4 targets, one at the center of each of the circles. For each reach from a starting point on the edge of a circle to the center, we measure the fraction of neural activity (A) or muscle activity (B) that a given unit contributes to the total at 50 ms intervals, and color the corresponding point along the straight line from the starting point on the outside edge to the target in the middle accordingly.

We see that the activation field for a given neuron in our model can vary quite dramatically depending on the three factors described above. For example, unit 3 is active early in the reach (indicated by the lighter colored ring around the edges of most of its circles), while unit 4 is active later in the reach for reaches above the head (indicated by the light colored centers of the top two circles) but earlier in the reach for reaches at shoulder level or below.

In Figure 3 B), we plot the fraction of muscle input that is contributed by each unit for each of the 4 muscles, as measured by the squared fraction of Dx that comes from each unit, divided by the sum of squares from all the units. The sign indicates whether the neuron contributes positively (thus activating the muscle) or negatively (thus deactivating it) at any given time point. The overall plotting method remains the same as in A). We again see that the contribution, this time to muscle input, can vary significantly over time, region, and reach direction. For example, with unit 3 we see an interesting biphasic change from activation of the shoulder extensor early in the reach to deactivation towards the end of the reach for reaches below the shoulder, while the pattern is reversed in reaches above the shoulder. Unit 4 has an even more complex pattern in the shoulder flexor of triphasic activation, then deactivation, followed by activation again, and only for upwards reach directions above the head.

IV. CONCLUSIONS

In this paper, we have shown that a relatively simple model of motor cortex dynamics is capable of controlling an arm via attractor dynamics that naturally arise through optimization. Further, our simulated motor populations revealed complex temporal and spatial activation fields across the entire reach space, something not seen in typical experiments in which all movement is done within a small task region. We believe that further study of these patterns may help us understand how the motor cortex organizes space over different scales.

REFERENCES

[1] Omrani M, et al. “Perspectives on Classical Controversies about the Motor Cortex.” Journal of Neurophysiology, vol. 118(3): 1828–48, 2017.

[2] Moran DW, Schwartz AB. Motor cortical representation of speed and direction during reaching. J Neurophysiol 82: 2676–2692, 1999.

[3] Morrow MM, Jordan LR, Miller LE. Direct comparison of the task-dependent discharge of M1 in hand space and muscle space. J Neurophysiol 97: 1786–1798, 2007.

[4] Cheney PD, Fetz EE. Functional classes of primate corticomotoneuronal cells and their relation to active force. J Neurophysiol 44: 773–791, 1980.

[5] Capaday C, et al. On the functional organization and operational principles of the motor cortex. Front Neural Circuits 7: 66, 2013.

[6] Park MC, Belhaj-Saif A, Cheney PD. Properties of primary motor cortex output to forelimb muscles in rhesus macaques. J Neurophysiol 92: 2968– 2984, 2004.

[7] McMorland JC, et al. “A Neuroanatomical Framework for Upper Limb Synergies after Stroke.” Frontiers in Human Neuroscience. 9: 2015.

[8] Vyas S, et al. Computation Through Neural Dynamics. Annu. Rev. Neurosci. 43: 249-275, 2015.

[9] Saxena S, Sarma SV, Dahleh M. Performance Limitations in Sensorimotor Control: Trade-Offs Between Neural Computation and Accuracy in Tracking Fast Movements. Neural Comput. 32: 5, 865-886, 2020.

[10] D’Aleo R, et al. Cortico-Cortical Drive in a Coupled Premotor-Primary Motor Cortex Dynamical System. https://ssrn.com/abstract=3859650.

[11] Agarwal R, et al. PMv Neuronal Firing May Be Driven by a Movement Command Trajectory within Multidimensional Gaussian Fields. J. Neurosci. 35(25): 9508-9525, 2015.

[12] Cheng EJ, Scott SH. Morphometry of Macaca mulatta Forelimb I. Shoulder and Elbow Muscles and Segment Inertial Parameters. J. Morphology. 245: 206-224, 2000.

[13] Trainin E, et al. “Explaining Patterns of Neural Activity in the Primary Motor Cortex Using Spinal Cord and Limb Biomechanics Models.” J Neurophysiol. 97: 15, 2007.

[14] Lillicrap TP, Scott SH. Preference distributions of primary motor cortex neurons reflect control solutions optimized for limb biomechanics. Neuron 77: 168–179, 2013.

[15] Pontryagin LS, et al. The mathematical theory of optimal processes. 1962.