

MODELING TARGET PRESENTATION TO THE MOTOR CORTEX USING
RECURRENT NEURAL NETWORKS

By

Yassin Fahmy

Copyright © Yassin Fahmy 2022

A Thesis Submitted to the Faculty of the
DEPARTMENT OF BIOMEDICAL ENGINEERING
In Partial Fulfillment of the Requirements
For the Degree of
MASTER OF SCIENCE
In the Graduate College
THE UNIVERSITY OF ARIZONA

2022

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Master's Committee, we certify that we have read the thesis prepared by *Yassin Fahmy*, titled *Modeling Target Presentation to the Motor Cortex Using Recurrent Neural Networks* and recommend that it be accepted as fulfilling the dissertation requirement for the Master's Degree.



Andrew Fuglevand, PhD

Date: May 9, 2022



Stephen Cowen, PhD

Date: May 9, 2022

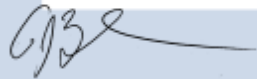


Jean-Marc Fellous, PhD


Date: May 10, 2022

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to the Graduate College.

I hereby certify that I have read this thesis prepared under my direction and recommend that it be accepted as fulfilling the Master's requirement.



Andrew Fuglevand, PhD
Master's Thesis Committee Chair
Biomedical Engineering

Date: May 9, 2022 

ARIZONA

Table of Contents

Abstract.....	4
Introduction.....	5
Methods.....	9
Results.....	20
Discussion.....	37
References.....	40

Abstract:

The introduction of multielectrode arrays capable of recording from hundreds of neurons simultaneously in the motor cortex (and other areas) has energized the field of motor control neurophysiology. The information obtained with such recordings, however, presents challenges with interpreting the functional significance of the simultaneous activities of large numbers of neurons during voluntary movement. One effective means to address this problem is to apply a dynamical systems approach, wherein population activity is considered to traverse a state space with each axis of the space represented by the firing of a single neuron. Despite the high dimensionality, neural trajectories in the motor cortex typically occupy only a low-dimensional subspace. One consistent observation based on recordings in monkey motor cortex during reaching movements is that the subspaces during movement preparation and movement execution are different from one another. An unanswered (and largely unasked) question has to do with the necessity and purpose of these different subspaces. To address this question, I modeled neural activity in the motor cortex using a recurrent neural network (RNN). The inherent structure of RNNs is like that of recurrently connected motor cortical neurons. I trained two identical RNNs on a delayed reaching task to eight targets distributed in a circle ('center-out-task'). The inputs were target location, timing of target presentation and go cues. Outputs were 'hand' kinematics representing movements to the targets. The two RNNs differed only in how target location was presented to the network; in one case the target signal was extinguished upon movement onset ('Interrupted input RNN'), while in the other, the target signal remained on ('Sustained input RNN'). In general, the activity patterns of individual neurons in the RNNs showed similarities to those recorded in the monkey motor cortex during this task. Interestingly, neural dynamics in the interrupted input RNN exhibited a similar orthogonalization of preparatory and movement related subspaces. However, for the sustained input RNN, preparatory and movement subspaces were practically the same. This finding

suggests that in the biological system, gating of sensory feedback about target location occurs near the time of movement onset, leading to a different representation in neural state space.

1. Introduction:

Motor preparation is typically used to describe all events occurring in the motor cortex approximately 40 ms prior to movement onset. Recordings of neurons in the motor cortex show distinct patterns of activity during motor preparation. Motor preparation and motor execution have been extensively studied and there exists several competing views regarding the role of the motor cortex in both events. Initially it was thought that motor cortical neurons' activity in the motor preparation epoch is simply a subthreshold activity of the corresponding movement neuronal activity (Tanji and Evarts 1976; Churchland et al., 2010; Churchland et al., 2014). Hypothetically, this would allow the motor cortical neurons' firing rates to come close to a threshold so that they can quickly produce a movement by simply raising their firing rates above the threshold needed to activate spinal motor neurons (Tanji and Evarts 1976; Churchland et al., 2010). However, the work of Churchland et al. (2010) showed that recordings of motor cortex neurons in monkeys display completely unrelated tuning between the preparatory epoch and movement epoch activity. Moreover, they also found that while the neural preparatory activity is usually stable in the preparatory epoch, activity during movement execution is multiphasic. This challenges the view that preparatory activity represents a subthreshold form of the activity associated with movement execution.

Multiple competing views exist regarding what (if any) behavioral parameters are represented in the firing of motor cortical neurons during voluntary movements. Originally motor cortical neurons were thought to be encoding distinct parameters of the upcoming movements, such as lower-level parameters like the pattern of EMG activity required to drive the appropriate muscles for the intended movement (Shenoy et al., 2013). Alternatively, the motor

cortex could encode higher-level parameters such as the desired kinematics of the upcoming movement (Georgopoulos et al., 1988 and Schwartz et al., 1999). These views of motor cortical function are referred to as ‘representational’ – namely, the activities of individual neurons are thought to represent some distinct aspect of movement.

In recent years major advances in neural recording technology have made it possible to record from hundreds of neurons simultaneously during a variety of behavioral tasks. When examining the activities of large populations of neurons made in such recordings, relatively few exhibited clear representational characteristics. A different way to analyze the activity of populations of motor cortical neurons is with a dynamical systems approach. In this approach, the collective activity of a population of interconnected neurons is thought to traverse a state space with each axis of this space representing the firing rate of one neuron. Thus, the number of dimensions in this space is directly related to the number of neurons in the recorded population (Gallego et al., 2018). However, such analyses revealed that neural activity in the motor cortex is confined to a substantially lower dimensional subspace (Churchland et al., 2014).

It appears that the internal dynamics generated by the motor cortical neurons’ recurrent connectivity, combined with an external sensory input that relays information about the external environment and limb position relative to the subject, generates consistent neural trajectories in low-dimensional state space. Interestingly, the emergent neural dynamics appear to occupy two separate subspaces: one associated with the preparatory period, and another for the movement epoch (Elsayed et al., 2016). A fundamental question prompted by this finding has to do with the meaning and necessity of separate subspaces associated with movement preparation and execution. Addressing this question experimentally, however, would be difficult. Therefore, here I address this question using a modeling approach involving recurrent neural networks (RNNs). The use of recurrent neural networks to model the activity of the motor cortex was first

done by Sussillo et al. (2015) and since then several other groups have used RNNs to successfully model the motor cortex dynamics (e.g. Michaels et al. 2016; Kao et al. 2019).

Since by design, the artificial neurons of RNNs are recurrently connected, the use of RNNs to model computations performed by the motor cortex during a specific behavioral task has been extensively used (Sussillo et al., 2015, Michaels et al., 2016, Kao et al., 2019). To train an RNN, a set of input and output time-varying signals are presented to the network. Then, through several training iterations, the network learns by adjusting its input, output, and recurrent (synaptic) weights in order to consistently produce an output that replicates the data the network was trained on (Hennequin et al., 2014, Sussillo et al., 2015, Michaels et al., 2016, kao et al., 2019). The activities of the network's artificial neurons can be considered a simulation of the operation of neurons in a biological network, in this case the motor cortex. Since RNNs are flexible in terms of training approaches, there exists multiple hyperparameters, regularization parameters, input and output designs, and even optimizers that can be used to train the network. Indeed, many input and output configurations have been used to model the same motor task (e.g. Sussillo et al. 2015; Michaels et al., 2016; Kao et al. 2019). The flexibility in RNN design and training to reproduce neural dynamics that resemble real neural data under different conditions provides insight into the actual organization of the motor cortex. Since we know the dynamical features of real neural data recorded in the motor cortex, we can use the flexibility of RNNs in terms of their design structure to test or generate hypotheses by comparing the dynamical features of a trained RNN to observed real dynamics.

As mentioned above, it appears that preparatory and movement computations occur in orthogonal dimensions (Elsayed et al. 2016). The reason for the orthogonality has not been fully explained. Specifically, what might trigger this type of organization? One possibility is that presentation and extinguishment of the targets to the motor cortex play roles in the abrupt change

of the motor cortical neurons' behavior between the preparatory and movement epochs. To investigate this possibility, I designed two identical RNNs in terms of having the same number of neurons, regularization, training parameters and output structure. Both RNNs were trained to perform a delayed reach task to one of 8 target locations spaced 45 degrees apart in a circle ('Center out task'). The two RNNs only differed in the duration of time the target was presented to the network. For the first RNN, the input signal carrying target location information was turned off at the time of the go cue, just before movement onset ('Interrupted input RNN'). For the second RNN, the input signal representing the target location persisted until the task was completed. Both networks were able to reproduce the x and y velocity vectors that complete the task. I found that both RNNs' artificial neurons show representational features in which individual neurons appear to have a preferred directional tuning. However, I found that only the interrupted RNN's dynamics showed features of population level dynamics consistent with real neural data observed in the motor cortex. Surprisingly, little orthogonalization between preparatory and movement spaces was evident when the input carrying the target location was maintained throughout the simulation. This finding suggests that internal gating of sensory inputs specifying target location during real reaching might underlie the shift between preparatory and movement states observed experimentally.

2. Methods:

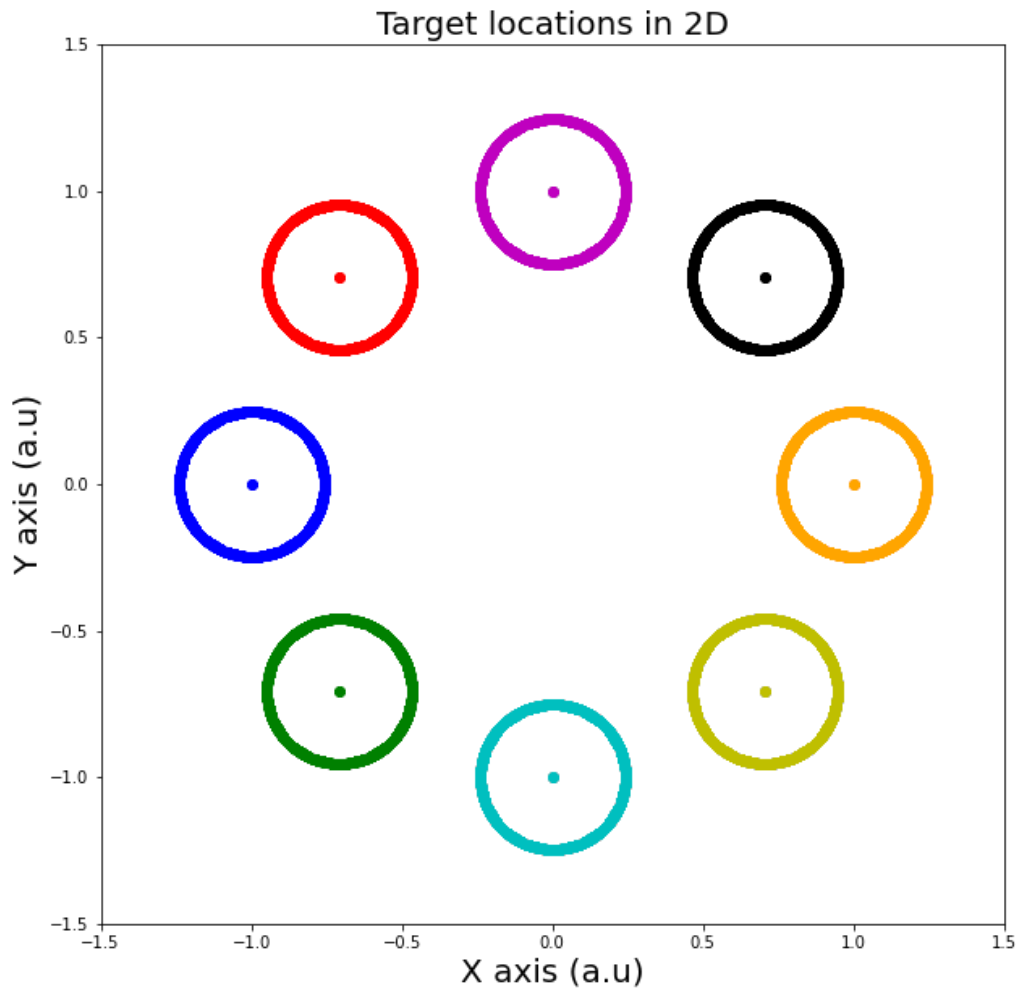


Figure 1. Target locations used to train both RNN models.

2.1 Task and data structure:

The training data used were collected from human subjects trained to perform the delayed reach task to one of 8 target locations spaced 45 degrees apart that collectively form a circle around the center ('center out task') as shown in Figure 1. A typical trial started with the subject holding a long (30 cm) joystick that controlled a cursor in the center position of the screen. After 200 ms, a 2.4 degrees (measured in visual angle with the subject seated approximately 60cm away from the screen) wide circular target appeared. The subject was instructed to hold the center position for a set amount of time of 1000 ms after which a go cue (a white ring around the target) signaled that the subject was free to execute the movement as fast as possible. The

trajectory of the cursor in the 2D plane was tracked as a function of time. This was measured as the continuous visual angle (in degrees) the evolving trajectory makes with the subject seated approximately 60cm away from the screen in the x and y planes. The target locations and trajectories were normalized before they were used to train the models, hence I refer to the units of target locations and trajectories as arbitrary units. The derivative of the normalized trajectory function was used to calculate the corresponding x and y velocities in arbitrary units as well (Michaels et al. 2016; Kao. et al. 2019). The 2D velocity signals were then used as the output signal the RNNs were trained on as shown in Figure 2.

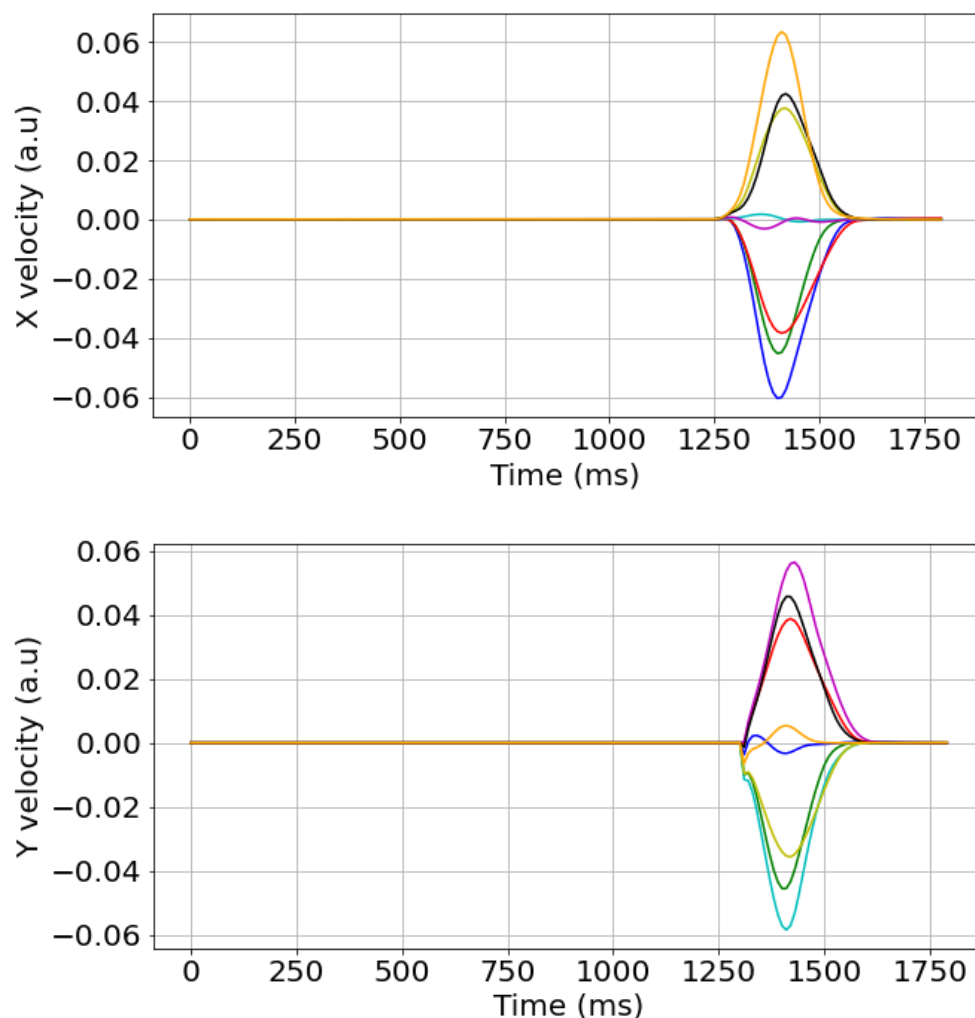


Figure 2. The output data structure used to train both RNN models. Colors indicate movements to different targets (see Fig. 1)

The input was composed of 3 signals. Two channels represented the x and y coordinates of the center of the target location in cartesian coordinates as a function of time as shown in the top two plots of Figures 3 and 4. The third input channel (bottom traces in Figures 3 and 4) represented the go cue signal which was first held at a value of one and then was set to zero (Sussillo et al., 2015; Michaels et al. 2016; Kao. et al. 2019). The only difference between the two RNNs was that the x and y position signals indicating target location were extinguished at the onset of the go cue for the interrupted RNN (Figure 3), and were left on for the sustained RNN (Figure 4).

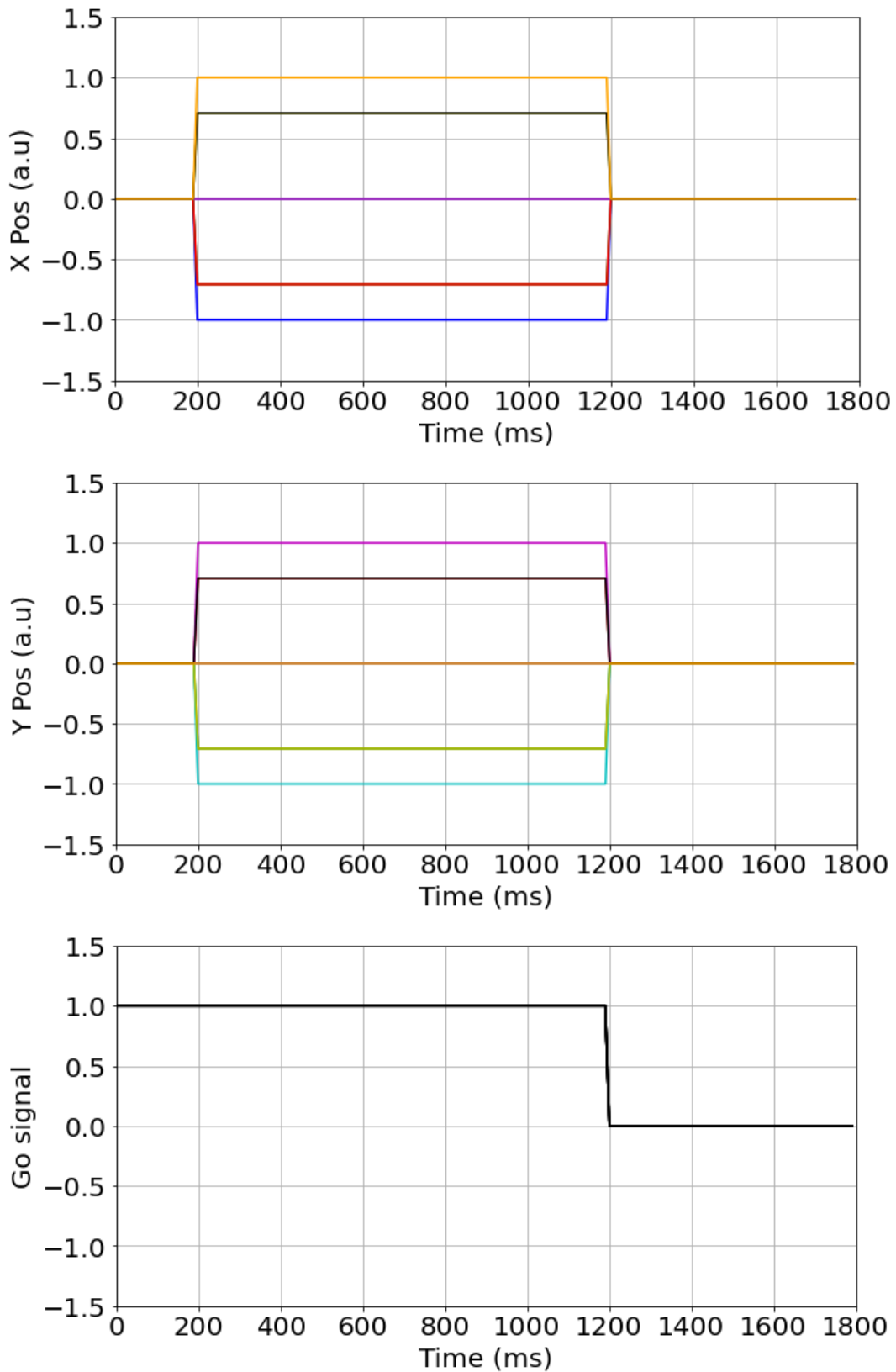


Figure 3. The input data structure to the interrupted RNN model. Each color indicates a different movement direction (see Fig. 1).

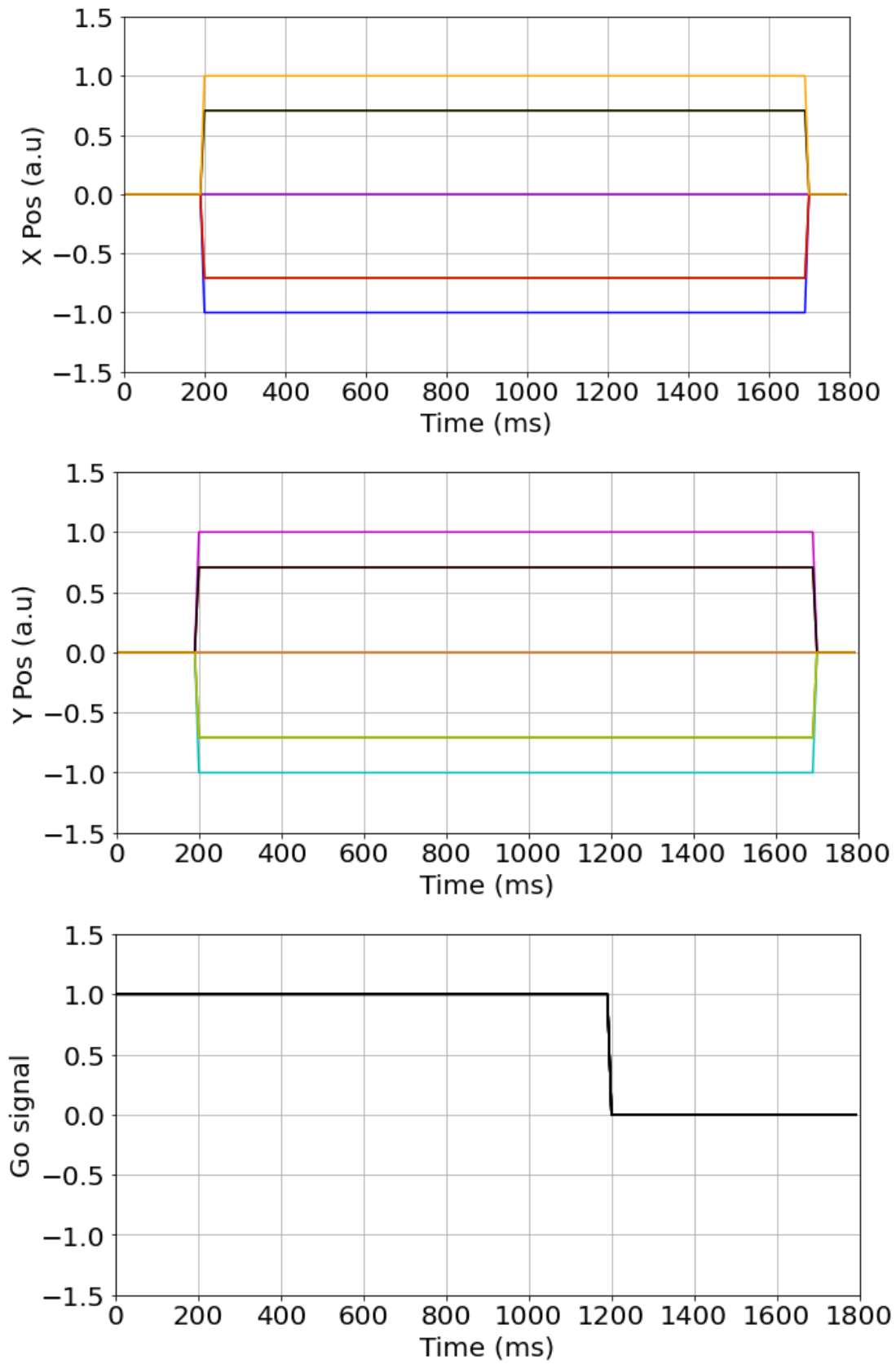


Figure 4. The input data structure to the sustained RNN model.

The dataset of human trials performing this task was composed of 1591 trials from one male subject. It is worthy to note that in the human trials the target was never extinguished for the entire length of a given trial. However, the interrupted RNN's input structure was extinguished by simulation. All trajectories were averaged per condition. The final input structure to the RNNs was composed of 8 reach directions, with 3 different simulated go cue times for a total of 24 conditions. I simulated a fixed reaction time of 100 ms after the Go cue as RNNs cannot model reaction times (Kao et al., 2019).

2.2 RNN description and model training:

Both RNNs were composed of 200 artificial neurons that receive $N_{in}=3$ time varying input signals and produce $N_{out}=2$ time varying output signals. The RNN produces a hidden state $\mathbf{x}_i(t)$ where the i th element describes the “current” in the i th artificial neuron. $\mathbf{f}(\mathbf{x}(t))$ describes the **tanh**($\mathbf{x}(t)$) function that transforms the network state to the artificial neuron firing rate $\mathbf{r}(t)$.

$$\mathbf{r}(t) = \mathbf{f}(\mathbf{x}(t))$$

The change in the network state (the firing rate) is summarized by the following equation described in Sussillo et al. (2015) and Kao et al. (2019):

$$\tau \frac{dx}{dt} = -x_i(t) + W_{rec} r_i(t) + W_{in} u(t) + b_{rec}$$

Where W_{rec} represents the recurrent weights matrix that connects the artificial neurons. W_{in} represents the input weights matrix that maps the input $u(t)$ onto each artificial neuron. b_{rec} is an added bias input channel with its own weights to help the network generalize. τ is a biological constraint that accounts for neural state decay to limit the magnitude of $\frac{dx}{dt}$. This value places a limit on the change in the network state that propagates from one time step to the next (Sussillo et al. 2015; Michaels et al. 2016; Kao. et al. 2019). It is important to note that the neurons' firing rate is determined by the **tanh** function which behaves linearly for input values between 1 and -1. Hence in this simulation we encounter negative firing rates typically not seen in the biological system, therefore I refer to the neurons' firing rates units as arbitrary (Sussillo et al. 2015;

Michaels et al. 2016; Kao. et al. 2019). The negative firing rates can simply be regarded as a decrease in the neurons' firing rate below its baseline activity.

The network output $z(t)$ is given by a linear readout of the matrix multiplication of the output weights matrix W_{out} and $\mathbf{r}(t)$.

$$z(t) = w_{out} \mathbf{r}(t)$$

The error function used to train the RNNs is the average squared error between the network output $z(t)$ and the two x and y velocity vectors. I used regularization parameters to mimic biological constraints in order to encourage the network to reproduce firing patterns similar to those seen in the motor cortex (Sussillo et al. 2015; Michaels et al. 2016; Kao et al. 2019). Specifically, I used a standard L2 regularization penalty on the input weights matrix, the output weights matrix, and the recurrent weights matrix. Essentially each of these regularization parameters adds an individual penalty term to the calculated error. This added term is composed of the $\lambda_{\text{Weight matrix}}$ (the tunable regularization parameter) multiplied by the squared sum of the weight matrix coefficients (Sussillo et al. 2015). The addition of these regularization parameters ensures that the RNN's neurons produce biologically plausible activity (Sussillo et al. 2015; Michaels et al. 2016; Kao et al. 2019).

I also used L2 regularization to limit the magnitude of the neurons' average firing rates across time in order to restrict the artificial neurons from large firing rates typically not encountered in a biological system (Sussillo et al. 2015; Michaels et al. 2016; Kao et al. 2019). This L2 regularization penalty was calculated as $\lambda_{\text{r}(t)}$ (the tunable regularization parameter) multiplied by R_{FR} . Where R_{FR} is the squared sum of all neurons' firing rates across all time for all conditions in one batch of training divided by the scalar multiplication of the total number of neurons by the number of time steps and number of conditions in a given batch (Sussillo et al. 2015).

I used the PsychRNN library that is publicly available for model building and training with the necessary modifications to fit the delayed reach center out task. Network training and simulations were run using Google colab for accelerated GPU based model training. The values used to train the model including hyperparameters and regularizations are reported in Table 1. Since the hyperparameter space is very large, I opted to use the same values reported in published papers that have previously used RNNs to model a delayed reach center out task of the same structure (Sussilo et al., 2015). Since the behavior is highly stereotyped, I let both RNNs train until they reached an R^2 value of 0.98.

Number of neurons	200
Time constant (τ)	50
Discretization bin width	10 ms
L_2 regularizer for W_{in} (λw_{in})	2×10^{-5}
L_2 regularizer for W_{out} (λw_{out})	2×10^{-5}
L_2 regularizer for W_{rec} (λw_{rec})	2×10^{-5}
L_2 regularizer for $r(t)$ ($\lambda r_{(t)}$)	1×10^{-3}
Activation function	$\tanh()$
Initial learning rate	1×10^{-4}
Max gradient norm	1

Table 1. Regularization and hyperparameter values used to train both RNN models.

2.3 Data analysis:

2.3.1 Cosine tuning:

To consider the possibility of representational activity in the modeled neurons, I used a modified version of the classic cosine tuning function previously reported by Georgopoulos et al. (1988) to fit neural activity associated with this task, as:

$$d_i(\mathbf{M}) = b_i + k_i \cos(\vartheta_{C_i \mathbf{M}})$$

where d_i is the average discharge rate of the i th neuron in the movement direction \mathbf{M} during the movement epoch. C_i describes the neuron's preferred direction. b_i and k_i are constants. $\vartheta_{C_i \mathbf{M}}$ describes the angle formed between C_i and \mathbf{M} . I report the average R^2 value for the fit of both populations of neurons in both RNNs.

All R^2 values reported can be considered modified R^2 values calculated using the following equation:

$$R^2 = 1 - \frac{RSS}{TSS}$$

Where RSS is the residual sum of squares calculated as the squared sum of the difference between the independent values and the predicted values. TSS is the total sum of squares calculated as the squared sum of the difference between the independent values and their mean. In this context if the model's predictions are close to the mean of the independent values, then $\frac{RSS}{TSS} = 1$ and the calculated R^2 would be zero. This makes it possible to have negative modified R^2 values.

2.3.2 Correlation between preparatory and peri-movement epochs:

Following Churchland et al. (2010), I computed the average firing rate during the preparatory epoch of each condition for each neuron. This resulted in a preparatory neural tuning 2D array (neuron x condition) where each neuron's tuning was described by a vector with one entry per condition for a total of 8 conditions. To evaluate the similarity in neural activity across preparatory and movement states, for each neuron I computed the correlation of its firing rates for the 8 conditions every 10 ms until 400 ms after movement onset with its average firing rates

per condition during the preparatory period. As described by Churchland et al. (2010), the mean correlation will be high if the neurons' preference remains similar across the preparatory and movement epochs.

2.3.3 Principal component analysis:

The predictions made by the network produces 2D velocity time varying signals which can be integrated over time to produce the movement trajectory. At each time point during a given movement, there existed a set of instantaneous firing rates, one for each of the 200 neurons in the RNN. Principal component analysis (PCA) was used as a dimensionality reduction method to visualize the networks' dynamics in a lower dimensional space (Lara et al., 2018). The firing rates of each artificial neuron were averaged across conditions (target location) to produce an array of 8 conditions for N neurons at each time step. The mean firing rate of each neuron across all conditions was calculated for each time step and then subtracted from each neurons' firing rate at each time step (Lara et al 2018). This preprocessing is done to allow PCA to capture the variance in dimensions where the responses are selective across conditions (Lara et al 2018). I performed PCA treating each neuron as a variable to obtain 4 preparatory principal components and 4 movement principal components that captured the maximum percentage of variance in the neural population response. The preparatory epoch was defined as a 300 ms period starting 400 ms before the go cue, while the movement epoch was defined as a 200ms epoch from the time of movement onset as described in (Elsayed et al., 2016, Lara et al 2018). The method used to find the angles between the preparatory dimensions and the movement dimensions is described by the following equation:

$$\theta = \cos^{-1} [(a \cdot b) / (|a| |b|)]$$

Where 'a' and 'b' are the two (200 element) vectors representing the principal axes found by PCA and θ is the angle between both vectors.

To quantify how much variance was captured by each principal component, I projected the movement and preparatory epoch neural population activity (separately) onto the preparatory principal components and I calculated the amount of neural variance that each of the preparatory principal components captures in each epoch. I repeat the same analysis for the movement principal components for both RNN models (Elsayed et al. 2016). I then projected the whole neural population response onto the first 2 preparatory principal components for visualization purposes in 2D space (Elsayed et al. 2016).

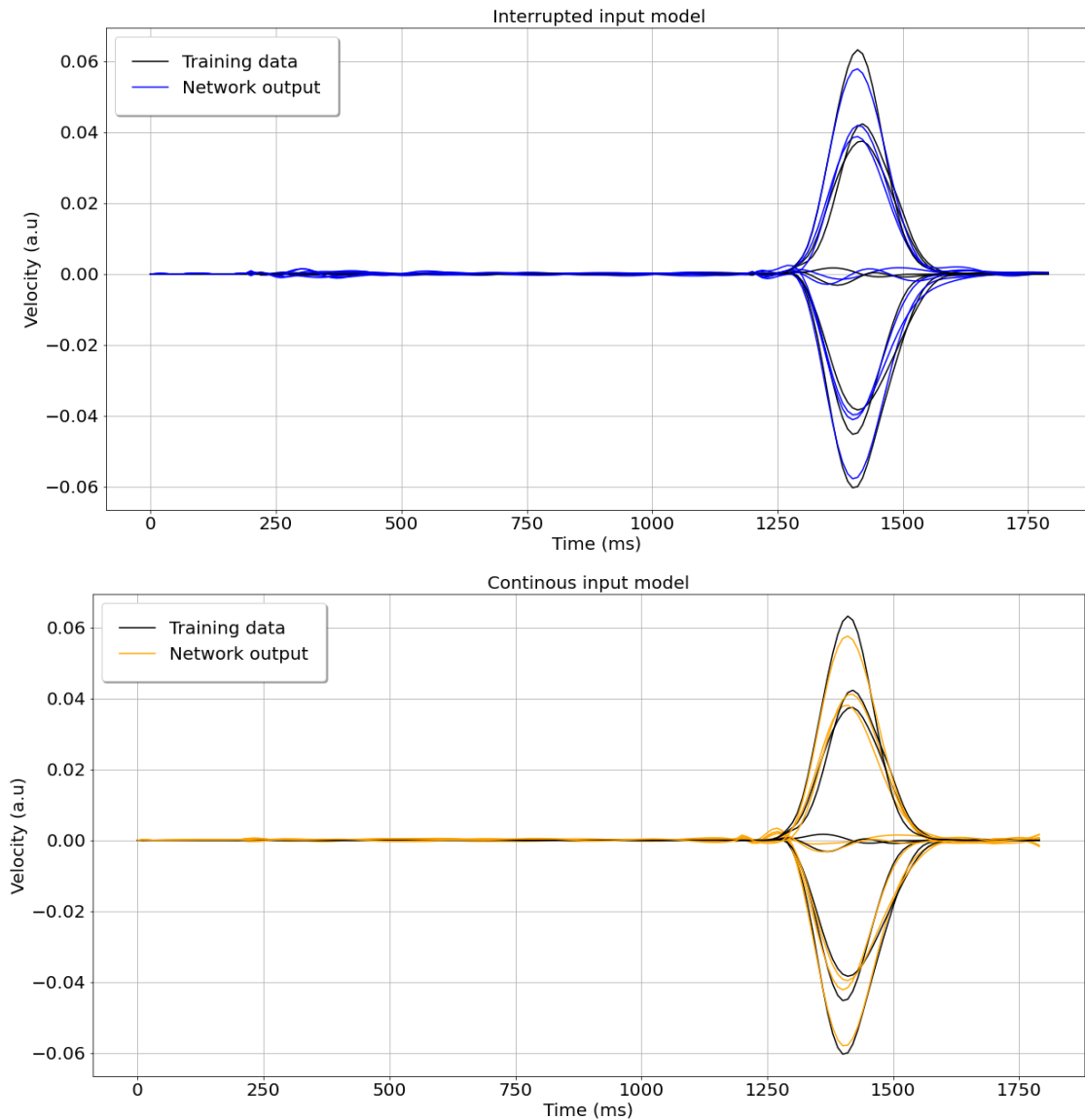


Figure 5. The trained RNNs’ x velocity output and the training data overlaid. Top panel shows the interrupted RNN’s output in blue. Bottom panel shows the sustained RNN’s output in orange.

3. Results:

3.1 Model performance

Both RNN models were able to learn the delayed reach center out task and produce x and y velocity vectors that matched the training data ($R^2 = 0.98$). Figure 5 shows the training data and the model output overlaid for both models. The top plot shows the interrupted RNN’s x-velocity output in blue while the sustained RNN’s velocity output is shown on the bottom plot in orange.

Both networks were able to hold during the variable length holding period and produce the corresponding velocity profile that successfully completes the task. By integrating the output velocity vectors of both networks separately one can recreate the trajectory of movement over time in cartesian coordinates. Figure 6 shows the plot of the x and y trajectories against each other in a 2D plot, the reach trajectories for each target (black circle) and the corresponding network output (blue for the interrupted RNN and orange for the sustained RNN). It is worthy to note that the network was never directly trained to produce the given trajectories. Rather the output signal consisted of x and y velocity vectors as a function of time. Nonetheless, the networks learned to adjust its synaptic weights to create a function capable of taking x and y coordinates and producing x and y velocity profiles that upon integration produce a trajectory in 2D space to the given x and y coordinates.

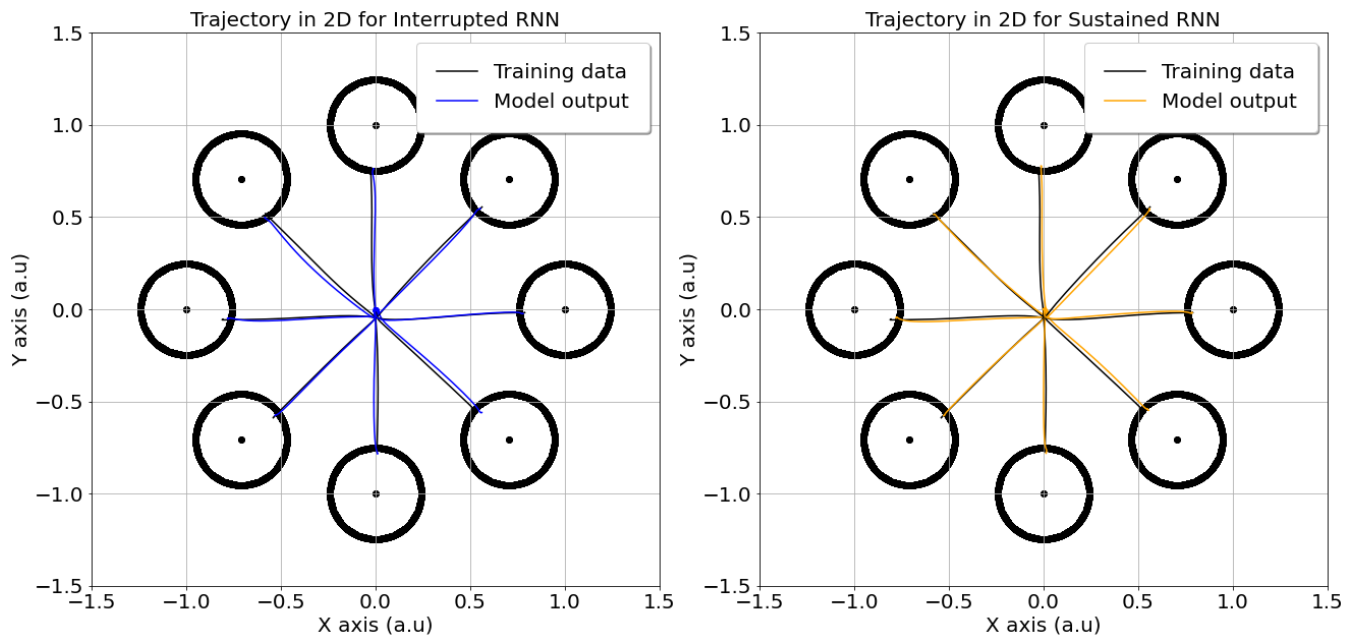
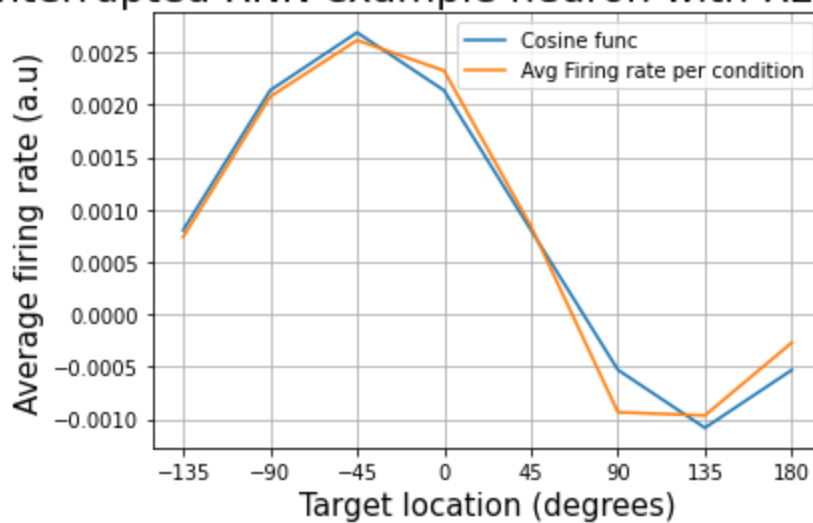


Figure 6. The trajectory formed by the integrated velocity outputs for both RNN models. The left panel shows the trajectory produced by the trained interrupted RNN. The right panel shows the trajectory produced by the trained sustained RNN.

3.2 Cosine Tuning in Modeled Neurons

Georgopoulos et al. (1988) and Schwartz et al. (1999) found that individual neurons of the motor cortex show strong directional tuning that can be fit to a cosine tuning function. Georgopoulos et al. (1988) showed that a neurons' discharge rate is highest for the direction they are tuned to and decreases gradually (resembling a cosine wave) in the directions immediately adjacent to the preferred direction. This property of real motor cortical neurons has been studied extensively and is a hallmark of the representational model of motor cortical computation. In order to contrast representational and neural dynamic approaches, I investigated whether artificial neurons show preferred directions as well. The average firing rate per condition for the movement epoch of each neuron was fitted to a cosine tuning function (see methods) with an average R^2 value of 0.91 for the goodness of fit of the interrupted RNN's neurons and a 0.93 R^2 for the Sustained RNN's neurons. Figure 7 shows an example neuron's average firing rate per condition and the cosine tuning function for the interrupted model (top plot) and sustained model (bottom plot). Over 90% of neurons of both models were strongly tuned to a preferred direction. Figure 8 displays a heat map of the average condition specific firing rates for 200 neurons plotted against the target location in polar coordinates for the interrupted RNN (top plot) and sustained RNN (bottom plot). The order of neurons was rearranged to showcase the strong tuning in the heat map plot. These results are consistent with Georgopoulos et al. (1988) findings in that a large population of the motor cortex neurons appear to have a preferred direction.

Interrupted RNN example neuron with R2 fit=0.98



Sustained RNN example neuron with R2 fit=1.0

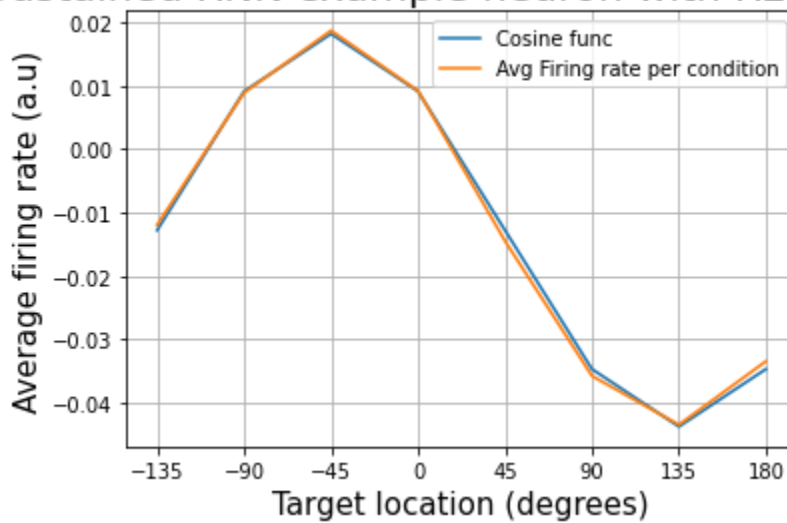


Figure 7. Two example neurons' cosine tuning fit from each of the RNN models. Top panel shows the average firing rate per condition for the interrupted RNN's example neuron. The bottom panel shows the average firing rate per condition for the sustained RNN's example neuron.

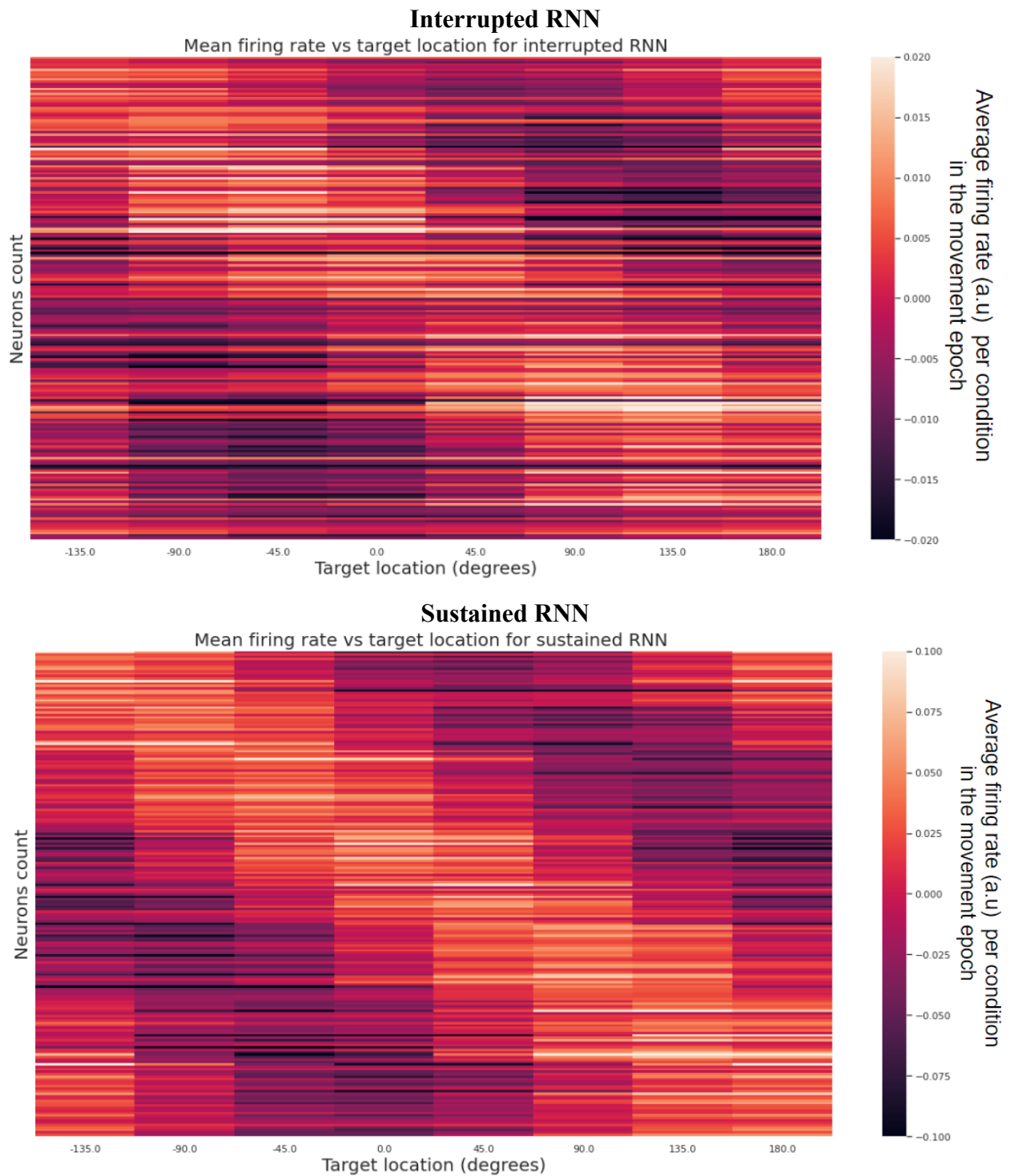


Figure 8. The average firing rate of the whole neural population for the 2 RNN models per condition (lighter colors indicate higher firing rates). Top panel shows the average firing rates per condition for the interrupted RNN. Bottom panel shows the average firing rate per conditions for the sustained RNN.

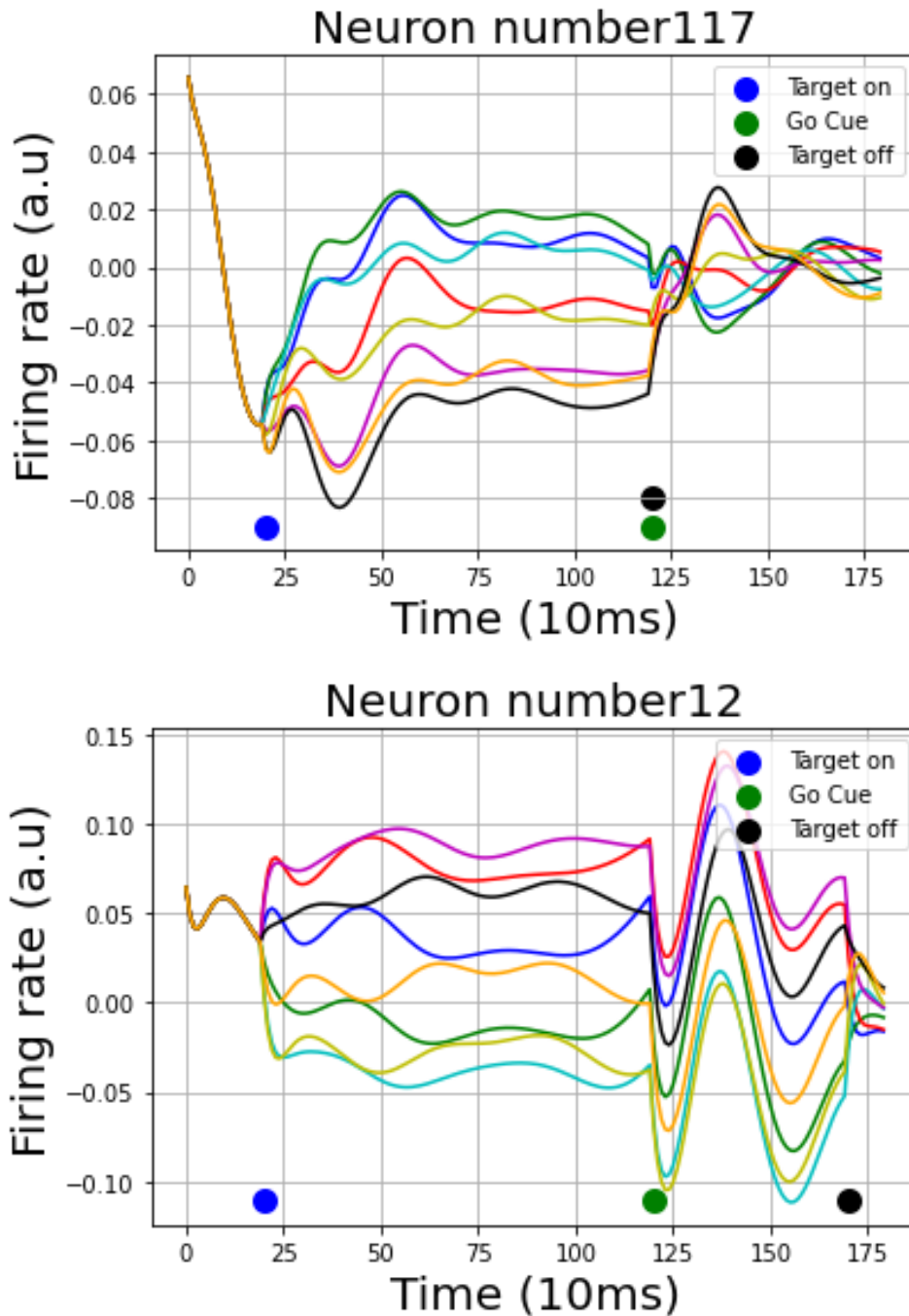


Figure 9. Example neurons' firing rate for both RNN models. Top panel shows the firing rate of neuron #117 of the interrupted RNN. The bottom panel shows the firing rate of neuron #12 of the sustained RNN.

3.3. Comparison of neural activity in the preparatory and movement periods

Recordings of real neurons in the motor and premotor cortex of macaque monkeys during a delayed reach task were found to display specific characteristics on the single neuron level

(Churchland et al., 2010). During the preparatory period, neurons show relatively stable condition specific tuning starting shortly after target onset and continuing up to around the time of movement onset. Near movement onset Churchland et al. (2010) found that neurons start to show multiphasic activity and a change in the preferred direction. Similarly, a RNN learns to hold the position at the center during the preparatory epoch until the go signal is given (the network is released), from which the network's internal dynamics produce the corresponding output. Figure 9 shows examples of a neurons' firing rates for the interrupted RNN (top plot) and another for the sustained RNN (bottom plot). Both neurons' firing rates qualitatively show stable condition specific tuning during the preparatory epoch and a transition to multiphasic activity at movement onset. Pre-target onset the activity of the two example neurons in Figure 9 would represent the neurons baseline activity; any increase or decrease in the neurons' firing rates post target presentation would simply describe a positive or negative deviation of the neurons' firing rates with reference to baseline activity. Upon target presentation Churchland et al. (2010) also found that motor cortical neurons recorded in 4 different monkeys showed near zero mean correlation between the preparatory epoch's tuning and that of the movement epoch. To compare the modeled neurons' stability of tuning across time I computed the correlation between the preparatory epoch's tuning and its peri-movement tuning at each time step for each neuron, replicating Churchland et al. (2010). Figure 10 shows the mean correlation across all 200 artificial neurons between the preparatory epochs condition specific tuning and each time step for both models. Time is centered with respect to movement onset in order to highlight the essential takeaway from this plot. The interrupted model's (shown in the blue trace) mean correlation drops to near zero at the time of movement onset resembling what has been observed in real data. However, the sustained model's mean correlation continues to stay relatively unchanged despite the onset of movement. It is worthy to note that both models were able to complete the task and produce the correct output (with 0.98 R^2) as was shown in Figure 5.

Qualitatively, it can be seen in Figure 9 (bottom panel) that the tuning for the example neuron from the sustained model was relatively well maintained across the whole trial despite the multiphasic activity seen during the movement epoch. On the other hand, tuning changes dramatically between the preparatory and movement periods for the example neuron from the interrupted RNN (Figure 9, top panel).

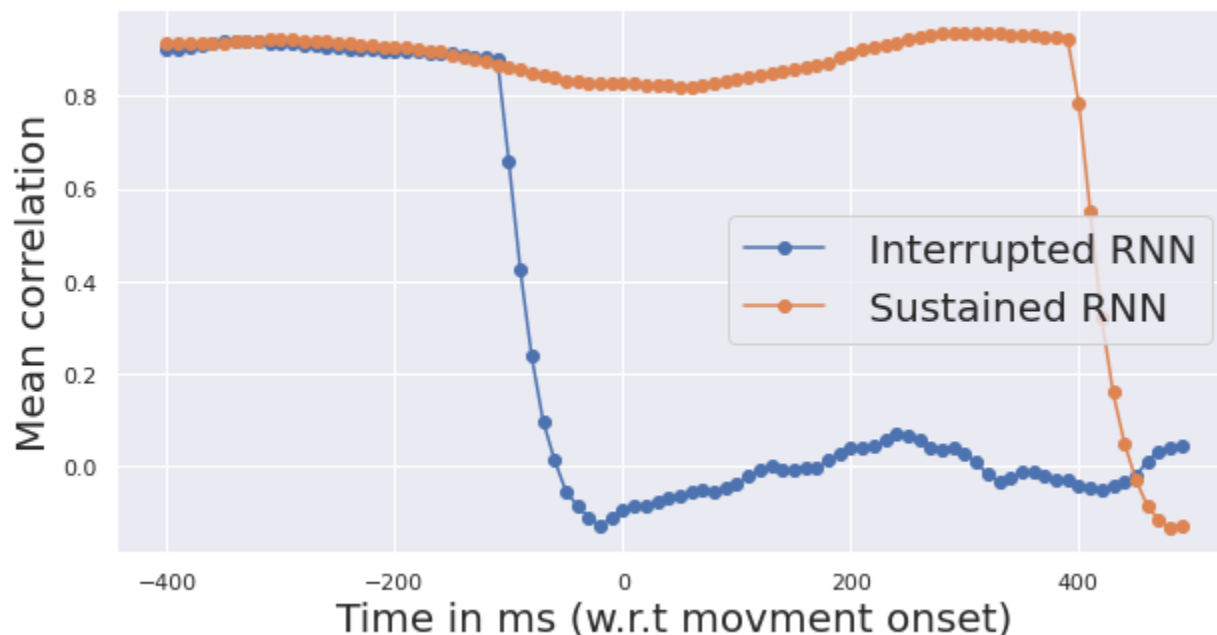


Figure 10. Correlation between preparatory and peri-movement tuning. Average correlation as a function of when perimovement was assessed. The interrupted RNN’s mean correlation is shown by the blue trace and the sustained RNN’s mean correlation is shown by the orange trace.

The persistence of preparatory condition specific tuning in the movement epoch in the case of the sustained RNN suggests that it might be utilizing an alternative dynamical computation to generate the corresponding output. One of the prominent features of the dynamical view is that the preparatory and movement epoch computations happen in two orthogonal sets of dimensions (Elsayed et al., 2016 and Vyas et al., 2020). This feature was first demonstrated by (Elsayed et al., 2016) by showing that the correlation matrix between the neurons’ firing rates in the preparatory epoch was completely different and unrelated to the

correlation matrix structure during the movement epoch. One can think of this as the same interconnected neurons being able to function as two distinct circuits. Figure 11 shows the correlation matrix plots for the 200 neurons of the interrupted RNN during the preparatory epoch (top left plot) and the movement epoch (top right plot). Structure in the correlation matrix in the preparatory period partially dissipates in the movement period. When plotting the correlation coefficients for the preparatory and movement epochs against one another, little inherent correlation was seen ($R^2 = -0.12$, Figure 11, bottom). This finding is consistent with that for motor cortical neurons (Elsayed et al., 2106). Figure 12 shows the same correlation plots for the sustained RNN where the scatter plot (the bottom plot) of the preparatory correlation values plotted against the corresponding movement epoch correlation values lie more or less along the diagonal (0.57 R^2 value). This suggests that the preparatory and movement dynamics generated by the sustained RNN share some of the same inherent structure.

Interrupted RNN

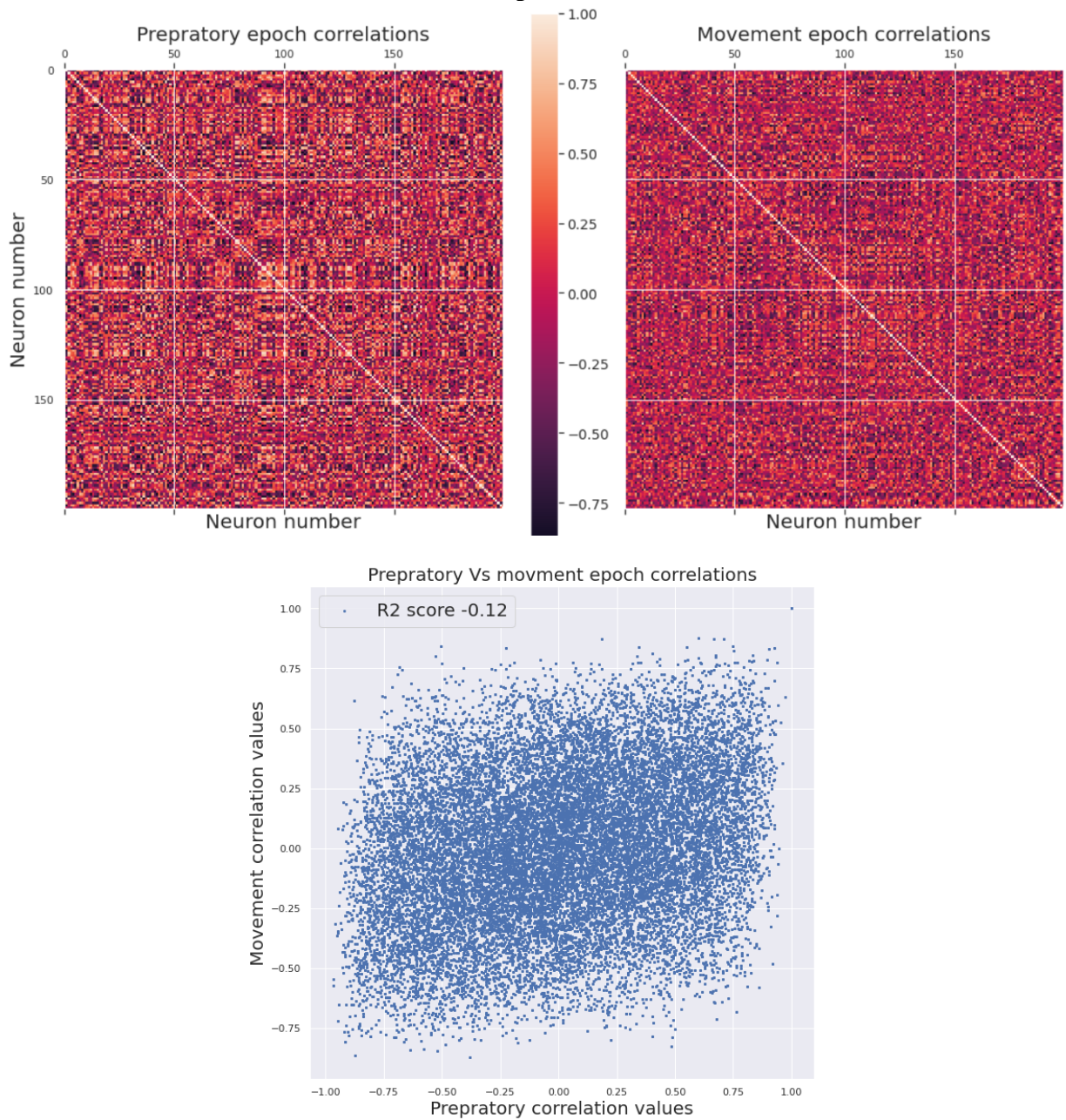


Figure 11. Preparatory and movement epoch correlation structure for all neurons in the interrupted RNN. Top row shows the preparatory correlation matrix on the left and the movement epoch correlation matrix on the right. The bottom plot shows the correlation for each neuron pair during the preparatory epoch plotted against the correlation for the same pair during the movement epoch.

Sustained RNN

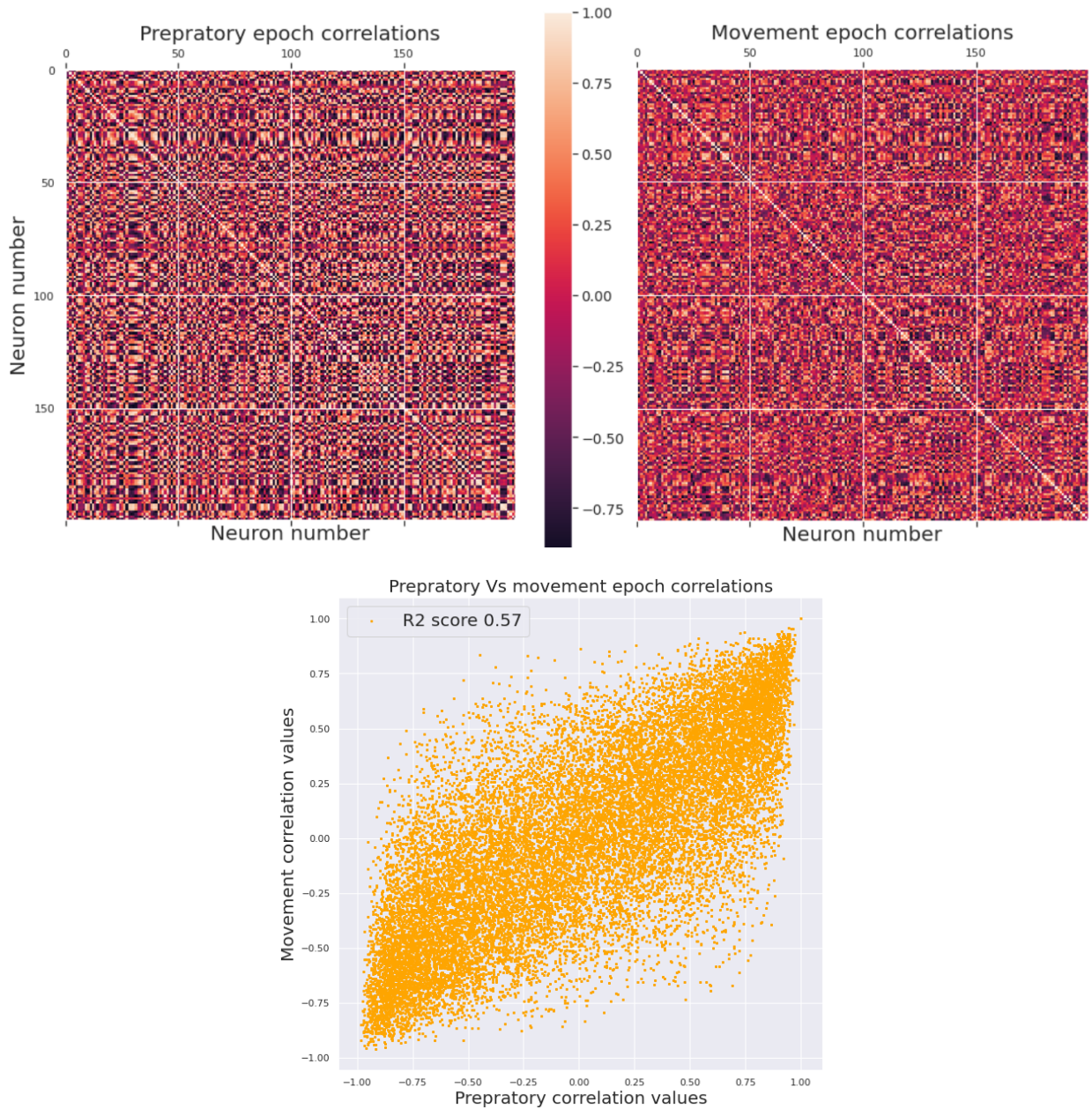


Figure 12. Preparatory and movement epoch correlation structure for all neurons in the sustained RNN. Top row shows the preparatory correlation matrix on the left and the movement epoch correlation matrix on the right. The bottom plot shows the correlation for each neuron pair during the preparatory epoch plotted against the correlation for the same pair during the movement epoch.

To this extent I perform PCA, a dimensionality reduction technique designed to identify the preparatory and movement dimensions that capture the highest percentage of variance in their corresponding epochs (Elsayed et al., 2016). I calculated how much variance each

preparatory and movement component captured from each epoch as shown in Figure 13. The top 4 preparatory components (top plot, red bars) captured a large percentage of the variance in the preparatory epoch. Likewise, the top 4 movement principal components captured the largest percentage of the movement epoch data variance (bottom plots, green bars). For the interrupted RNN (top and bottom left plots), the top 4 preparatory principal components capture very little movement epoch data variance and the top 4 movement principal components captured very little preparatory epoch data variance resembling the results observed in real data presented by (Elsayed et al., 2016 their figure 4). However, for the sustained RNN (top and bottom right plots), the top 2 preparatory principal components captured a large percentage of the variance in the movement epoch (top right plot, green bars) and the top 2 movement principal components captured a large percentage of the variance in the preparatory epoch (bottom left plot, red bars). This finding suggests that the sustained RNN seems to use a dynamical computation that involves sharing common dimensions between the preparatory and movement subspaces.

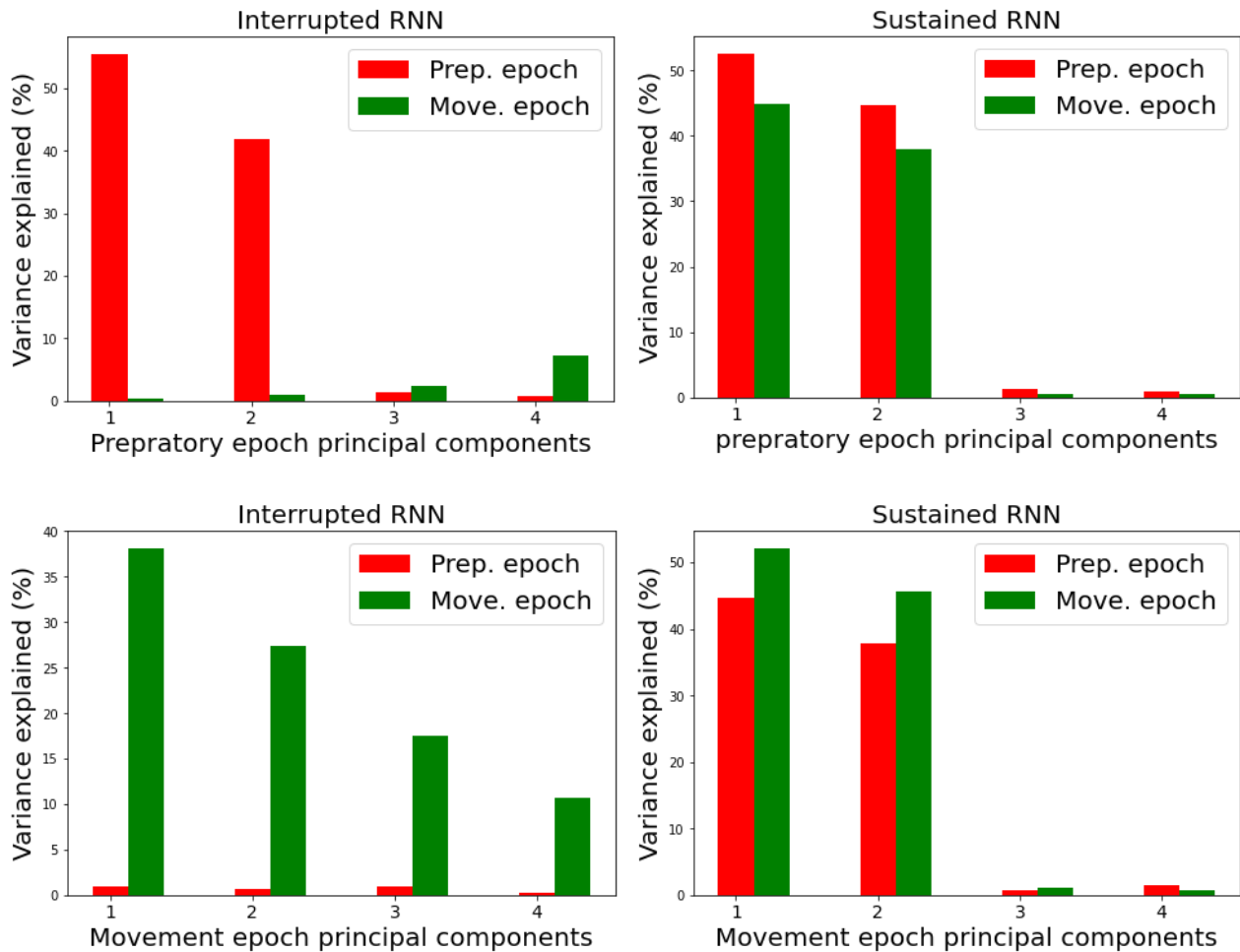


Figure 13. Percentage of variance explained by preparatory and movement principal components. Top plots show the percentage of preparatory epoch data variance (in red) and movement epoch data variance (in green) explained by the top 4 preparatory principal components for the interrupted RNN (left plot) and sustained RNN (right plot). Bottom plots show the same analysis but for the top 4 movement principal components.

To further quantify the relationship between the preparatory and movement subspaces for both RNNs, one can calculate the angle between the preparatory and movement dimensions represented by the principal axes identified previously by PCA (Figure 14). For the interrupted RNN (shown in blue) the angle between each 2 pairs of dimensions (one preparatory and movement dimension) ranges from 84 to 86 degrees. This is consistent with findings of inherent orthogonality between movement and preparatory dimensions in real data recorded from the motor cortex (Elsayed et al., 2016). However, for the sustained RNN the angle between the

preparatory and movement dimensions range from 29 to 90 degrees. This finding outlines that despite the sustained RNN's ability to complete the task with high accuracy and show features of the representational model such as a preferred direction tuning, the network fails to replicate key features of the dynamical model observed in real data. It is notable that the only difference in the sustained RNN as compared to the interrupted was that the target locations remained available during the entire preparatory and movement periods.

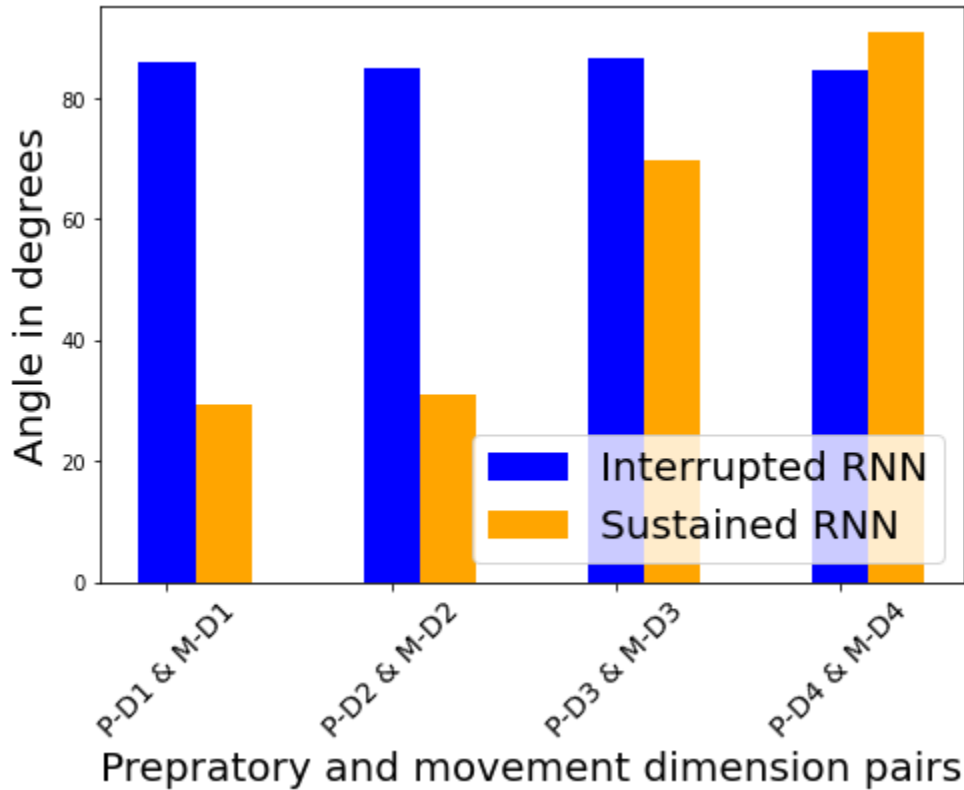


Figure 14. The angles between the pairs of principal axes for each RNN model separately, displayed side by side. ‘P-D1’ stands for preparatory dimension number 1. ‘M-D1’ stands for movement dimension number 1.

3.4 Neural Dynamics During Preparatory and Movement Periods

Elsayed et al. (2016) and Lara et al. (2018) showed that in the preparatory subspace, target presentation triggered neural trajectories to diverge from the center and to occupy states that resemble the spatial arrangement of target locations. They also showed that upon delivery of the go cue, the neural trajectories in the preparatory subspace collapsed back to the center (and

stayed at the center for the remainder of the trial) marking the end of the preparatory dynamics. On the other hand, movement dynamics start diverging in the movement subspace around the time of movement onset while there is virtually no activity in the preparatory dimensions. Given that the sustained RNN showed evidence of using the same dimensions for the preparatory and the movement epoch dynamics, I performed PCA to identify the preparatory subspace that contains the preparatory dynamics (Elsayed et al., 2016 and Lara et al., 2018) and to project the whole population response onto the preparatory subspace (Elsayed., et al 2016). Figure 15 shows the neural dynamics in the preparatory subspace for the 2 RNN models. The left top and bottom plots of Figure 15 show that upon target onset the neural trajectories diverge in preparatory state space for both RNN models (interrupted RNN shown in top plot in blue and sustained RNN shown in bottom plot in orange) to a spatial organization that is representative of the target locations. Upon delivery of the go cue, only the interrupted RNN's neural trajectories start to collapse back to the center (top middle plot, Figure 15) marking the end of preparatory dynamics and resembling what has been observed in real data. Meanwhile, the neural trajectory for the sustained RNN remained in the previously attained preparatory states (bottom middle plot, Figure 15). At the time of movement onset the interrupted RNN's neural trajectories are maintained at the origin (top right plot, Figure 15). Meanwhile the sustained RNN's neural trajectories remained in their previously attained preparatory states during movement generation (bottom right plot, Figure 15). The averaged Euclidean distance for the all neural trajectories relative to the initial state in the preparatory subspace across all conditions for both RNN models are shown in Figure 16. This demonstrates that while the interrupted RNN's (blue trace) neural trajectories collapse back to the origin around the time of movement onset, the sustained RNN's neural trajectory remains divergent throughout the movement epoch.

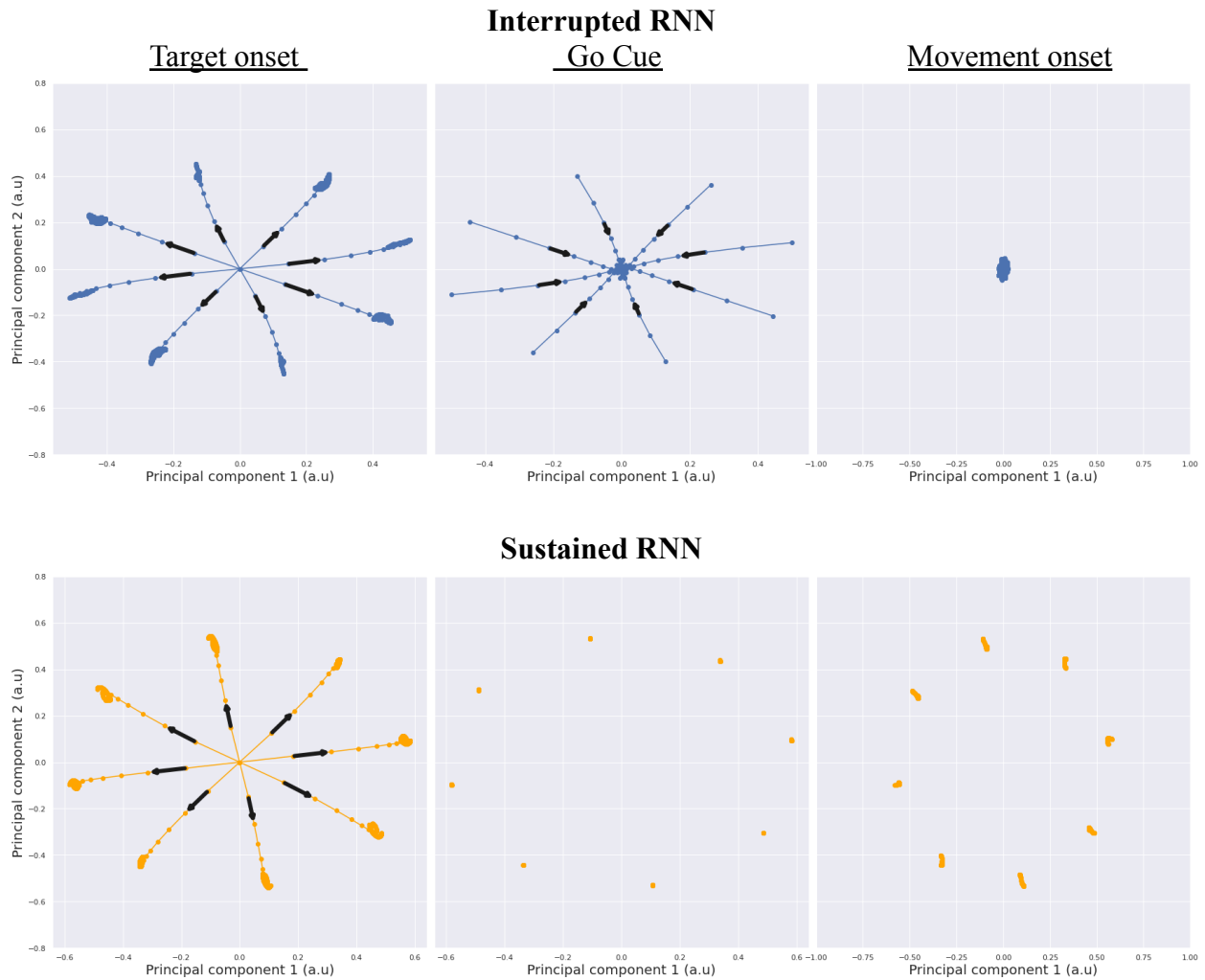


Figure 15. Activity in the preparatory subspace in response to key task events. Each trace corresponds to a different reach direction. The neural trajectory in response to target onset up to the time before the go cue is shown in the top left plot for the interrupted RNN and the bottom left plot for the sustained RNN. The top middle plot shows the neural trajectory from the time of go cue up to before the time of movement onset for the interrupted RNN and the bottom middle plot shows the same for the sustained RNN. The top and bottom right plots show the neural trajectory from the time of movement onset till the end of the trial for the interrupted and sustained RNNs respectively.

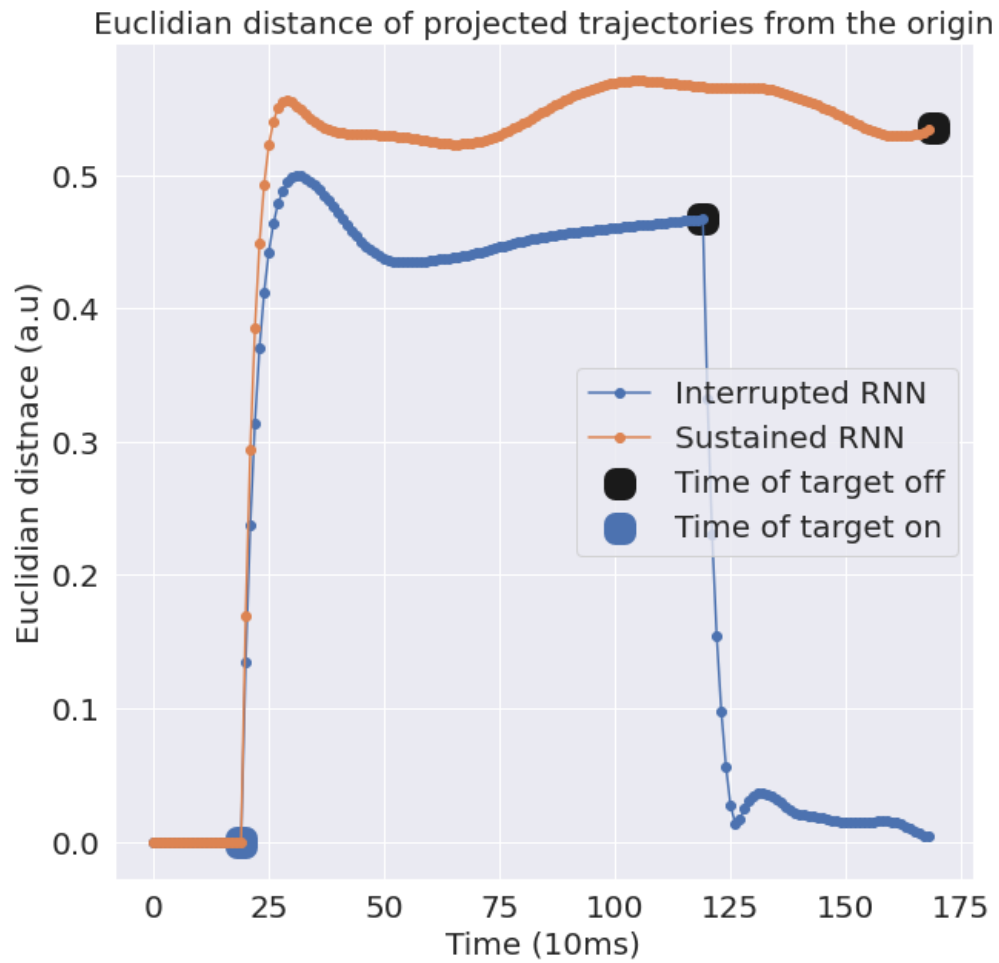


Figure 16. The average euclidean distance across all conditions for the entire length of the trial. Blue trace shows the euclidean distance for the interrupted RNN and the orange trace corresponds to the sustained RNN.

4. Discussion:

I have built and trained two identical RNNs on the delayed reaching center out task. Both models had the same number of neurons, same regularization hyper-parameters and same output structure. Both produced near perfect replication of output hand kinematic. The two models only differed in the time the input signal carrying the target location information was extinguished. For the interrupted RNN the target location signal was extinguished at the time of the go cue; meanwhile, the target location signal for the sustained RNN was left on until the trial was completed.

The training of both networks did not involve any cosine tuning restrictions for the RNNs' artificial neurons. Nonetheless, the majority of neurons in both RNN models showed a preferred direction fit well by a cosine function consistent with experimental findings of motor cortical neurons (Georgopoulos et al., 1988; Schwartz 1999).

I also found that the correlation in condition specific tuning over time differed between both RNN models. The interrupted RNN's neurons' condition specific tuning during the preparatory epoch was unrelated to their tuning during the peri-movement period, similar to what has been observed in real data from the motor cortex (Churchland et al., 2010). Conversely, the sustained RNN's neurons showed a strong correlation between the preparatory and movement epoch condition specific tuning. While both RNN models' neurons showed other features observed in real motor neurons such as stable tuning during the preparatory epoch and multiphasic activity during the movement epoch, it appears that the two RNNs used different computations to complete the same task with equal fidelity.

The change of neural tuning between the preparatory and movement epochs was found to be due to the nature of the dynamical computation performed by the neural population occupying two separate orthogonal subspaces for the preparatory and movement epochs (Kaufman et al., 2014 and Elsayed et al 2016). I found that only the interrupted RNN's neural population showed

unrelated correlation matrices for the preparatory and the movement epochs. Furthermore, upon analysis of the preparatory and movement principal components, only the interrupted RNN displayed the inherent orthogonality and separation of the preparatory and movement subspaces observed in real data from the motor cortex. On the other hand, the sustained RNN showed a high degree of overlap between the preparatory and movement subspaces.

Since both RNNs appear to be using different dynamics to successfully complete the same task, I opted to use principal component analysis to identify the preparatory subspace. In the motor cortex neural trajectories diverge to discrete preparatory states upon target presentation. Upon delivery of the go cue, neural trajectories start to collapse, marking the end of the preparatory dynamics and the hold period of the task. Furthermore, at the time of movement onset there is virtually no activity in the preparatory dimensions. As such, there is a separation of preparatory and movement dimensions during the preparatory and movement dynamical computations. The neural trajectory of the interrupted RNN in the preparatory subspace mimicked the preparatory dynamics recorded in the motor cortex (Elsayed et al., 2016 and Lara et al., 2018). However, the sustained RNN's neural trajectory failed to collapse upon delivery of the go cue.

Even though the dynamical characteristics of the population response in the motor cortex has been previously described (Churchland et al., 2010, Churchland et al., 2014, Michaels et al., 2016, Elsayed et al., 2016 and Lara et al., 2018), the trigger for the orthogonality of the preparatory and movement dimensions has not been characterized before. My analysis suggests that the interrupted RNN appears to be more consistent with the dynamical features observed in the motor cortex as shown by (Elsayed et al., 2016 and Lara et al., 2018) when compared to the sustained RNN. It is important to note that the real experimental paradigm in the analysis presented by (Elsayed et al., 2016 and Lara et al., 2018) did not involve turning the target off. The monkeys were able to see the target location during the entire length of the trial. Yet, when

replicating this paradigm in the sustained RNN by training the network on a target location input signal that remains on for the entire length of the trial, the emergent dynamics from the RNN did not resemble the real data. On the other hand, the interrupted RNN in which the target location signal was extinguished prior to movement onset (and therefore, unlike the paradigm used experimentally), showed dynamical features consistent with those observed in real data. These findings imply that target presentation to the motor cortex might be getting extinguished by some neural mechanism prior to movement onset. Kao et al. (2021) have recently shown modeling evidence that thalamocortical feedback loops could be gating sensory input about target location to the motor cortex coincident with extinguishing preparatory activity in the motor cortex. This is consistent with my findings that the timing of target deactivation is key in reproducing the real observed dynamical structure in the motor cortex.

Task input design has been shown previously to influence the dynamical computation made by RNNs (Kao et al., 2019). I recommend that future research considers investigation into thalamocortical feedback loops and their role in regulating the temporal target presentation to the motor cortex. The gating mechanism responsible for target presentation could possibly be modeled via a two-layer RNN; where the first layer would be composed of equally distributed excitatory and inhibitory neurons to act as a temporal on/off switch for target presentation to the subsequent layer modeling the motor cortex.

5. References:

Ames KC, Ryu SI, Shenoy KV. Neural dynamics of reaching following incorrect or absent motor preparation. *Neuron*. 2014 Jan 22;81(2):438-51. doi: 10.1016/j.neuron.2013.11.003. PMID: 24462104; PMCID: PMC3936035.

Churchland MM, Cunningham JP. A Dynamical Basis Set for Generating Reaches. *Cold Spring Harb Symp Quant Biol*. 2014;79:67-80. doi: 10.1101/sqb.2014.79.024703. Epub 2015 Apr 7. PMID: 25851506.

Churchland MM, Cunningham JP, Kaufman MT, Ryu SI, Shenoy KV. Cortical preparatory activity: representation of movement or first cog in a dynamical machine? *Neuron*. 2010 Nov 4;68(3):387-400. doi: 10.1016/j.neuron.2010.09.015. PMID: 21040842; PMCID: PMC2991102.

Elsayed GF, Lara AH, Kaufman MT, Churchland MM, Cunningham JP. Reorganization between preparatory and movement population responses in motor cortex. *Nat Commun*. 2016 Oct 27;7:13239. doi: 10.1038/ncomms13239. PMID: 27807345; PMCID: PMC5095296.

Gallego JA, Perich MG, Naufel SN, Ethier C, Solla SA, Miller LE. Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nat Commun*. 2018 Oct 12;9(1):4233. doi: 10.1038/s41467-018-06560-z. PMID: 30315158; PMCID: PMC6185944.

Georgopoulos AP, Kettner RE, Schwartz AB. Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of

movement by a neuronal population. J Neurosci. 1988 Aug;8(8):2928-37. doi: 10.1523/JNEUROSCI.08-08-02928.1988. PMID: 3411362; PMCID: PMC6569382.

Hennequin G, Vogels TP, Gerstner W. Optimal control of transient dynamics in balanced networks supports generation of complex movements. Neuron. 2014 Jun 18;82(6):1394-406. doi: 10.1016/j.neuron.2014.04.045. PMID: 24945778; PMCID: PMC6364799.

Kao JC. Considerations in using recurrent neural networks to probe neural dynamics. J Neurophysiol. 2019 Dec 1;122(6):2504-2521. doi: 10.1152/jn.00467.2018. Epub 2019 Oct 16. PMID: 31619125.

Kao TC, Sadabadi MS, Hennequin G. Optimal anticipatory control as a theory of motor preparation: A thalamo-cortical circuit model. Neuron. 2021 May 5;109(9):1567-1581.e12. doi: 10.1016/j.neuron.2021.03.009. Epub 2021 Mar 30. PMID: 33789082; PMCID: PMC8111422.

Kaufman MT, Churchland MM, Ryu SI, Shenoy KV. Cortical activity in the null space: permitting preparation without movement. Nat Neurosci. 2014 Mar;17(3):440-8. doi: 10.1038/nn.3643. Epub 2014 Feb 2. PMID: 24487233; PMCID: PMC3955357.

Lara AH, Elsayed GF, Zimnik AJ, Cunningham JP, Churchland MM. Conservation of preparatory neural events in monkey motor cortex regardless of how movement is initiated. Elife. 2018 Aug 22;7:e31826. doi: 10.7554/eLife.31826. PMID: 30132759; PMCID: PMC6112854.

Michaels JA, Dann B, Scherberger H. Neural Population Dynamics during Reaching Are Better Explained by a Dynamical System than Representational Tuning. PLoS Comput Biol. 2016 Nov 4;12(11):e1005175. doi: 10.1371/journal.pcbi.1005175. PMID: 27814352; PMCID: PMC5096671.

Moran DW, Schwartz AB. Motor cortical representation of speed and direction during reaching. J Neurophysiol. 1999 Nov;82(5):2676-92. doi: 10.1152/jn.1999.82.5.2676. PMID: 10561437.

Song HF, Yang GR, Wang XJ. Training Excitatory-Inhibitory Recurrent Neural Networks for Cognitive Tasks: A Simple and Flexible Framework. PLoS Comput Biol. 2016 Feb 29;12(2):e1004792. doi: 10.1371/journal.pcbi.1004792. PMID: 26928718; PMCID: PMC4771709.

Sussillo D, Churchland MM, Kaufman MT, Shenoy KV. A neural network that finds a naturalistic solution for the production of muscle activity. Nat Neurosci. 2015 Jul;18(7):1025-33. doi: 10.1038/nn.4042. Epub 2015 Jun 15. PMID: 26075643; PMCID: PMC5113297.

Tanji J, Evarts EV. Anticipatory activity of motor cortex neurons in relation to direction of an intended movement. J Neurophysiol. 1976 Sep;39(5):1062-8. doi: 10.1152/jn.1976.39.5.1062. PMID: 824409.

Vyas S, Golub MD, Sussillo D, Shenoy KV. Computation Through Neural Population

Dynamics. *Annu Rev Neurosci.* 2020 Jul 8;43:249-275. doi:

10.1146/annurev-neuro-092619-094115. PMID: 32640928; PMCID: PMC7402639.