Local network coordination supports neuroprosthetic control

William Liberti1, Xue Lily Gong2, Thomas Roseberry3, Rui M. Costa4 and Jose M. Carmena5, Fellow, IEEE

Abstract—Learning often involves adapting behavior in response to the inferred causes of success and failure. At the neural level, this can be the result of repeating activity patterns of neurons that lead to favorable outcomes. However, it is not clear how the contributions of individual cells to an ongoing behavior is assessed. Using a calcium imaging based closed loop Brain-Machine Interface (CaBMI), we trained mice to perform a neuroprosthetic task using the coordinated activity of a small ensemble of neurons in layer 2/3 of sensorimotor cortex. We find that that after an initial period of exploration, neurons that do not directly drive the effector decrease in variance and event frequency over the course of learning. However, a large fraction of these ‘indirect’ cells demonstrate robust spatiotemporal dynamics both before and after an animal achieves reward. Throughout a single 30 minute session, these spatiotemporal sequences increase in frequency and become more consistent. Our findings suggest that neuroprosthetic control is the result of an emergent, spatially organized network level solution, rather than the direct modulation of a few chosen output neurons.

I. INTRODUCTION

Several studies of learning have shown that spatiotemporal activity patterns of neurons which lead to desired behavioral outcomes are selected and consolidated [1]–[3]. The theoretical framework that describes this underlying computational problem is called ‘credit assignment’, and it is an important ingredient of reinforcement learning [4]. However, the single neuron rules that result in behavior modification or maintenance have been difficult to address, due to the immense fine-scale complexity of biological neural networks, the heterogeneity of natural behavior, and the lack of a causal relationship between an observed neuron’s activity and behavior.

To overcome these obstacles, we employ a reductionist approach where behavioral output is simply the optically recorded activity of a small ensemble of experimenter defined cortical neurons expressing genetically encoded fluorescent calcium indicators. This neuroprosthetic strategy uses closed-loop auditory feedback to guide a ‘fictive behavior’ where animals are rewarded when a specific network pattern is observed in a chosen neural population [5], [6]. Similar to natural motor tasks, previous work has shown that neuroprosthetic task performance tends to increases over days [5]. Surprisingly, BMI performance can increase even when the output cell identity is changed from day to day—suggesting that animals are learning to better assign reward credit to output neurons, rather than consolidating specific activity control patterns [7]–[9]. It is unclear how the local population of cells resolves the identity of the output cells on these short timescales. The remaining units that are not assigned a direct relationship with effector, called ‘indirect neurons’, have been shown to demonstrate tuning relative to BMI tasks, suggesting that some of these neurons may become incorporated into functional neuronal assemblies with ‘direct neurons’ [10], [11]. However, limited cell yields from chronic electrophysiology and an experimenter bias to assign the highest signal-to-noise units as output cells has provided limited opportunities to ask questions about the relative contributions of local networks [11].

We find that after an initial period of exploration, the local cortical neural population that does not directly drive effector output converges to form reproducible, spatiotemporally organized sequences of activity that increase in frequency and consistency throughout a 30 minute session [5], [10]. Taken together, our results suggest that neuroprosthetic control is achieved by increasing the frequency of a limited number of spatiotemporal network patterns, rather than the isolated modulation of individual cells.

II. METHODS

A. Animals and Surgical Procedure.

All animal procedures were performed in accordance with U.C. Berkeley IACUC regulations. Stereotactic surgery was performed in (n=5) Eμx1TRESCre or wild type mice as previously described [5]. 200-300nl of AA V2.9 CamKII.GCamp6f.WPRE.SV40 or AA V5-SYN-FLEX-GCaMP6f-WPRE-SV40 was used to preferentially label excitatory cells in cortex. Mice were given 4 weeks post-surgical recovery time to allow for sufficient protein expression.

B. Experimental Preparation.

In vivo imaging was performed at 30Hz using a commercial multi-photon microscope (Bruker, Ultima Investigator) driven by a mode-locked, tunable Ti:Sapphire laser (CHAMELEON ULTRA II) and set to 900-980 nm, steered via 8kHz resonant galvos through a 20x water immersion...
objective (XLUMPLFLN 20XW). Photons were collected with a GaAsP photomultiplier tube (Hamamatsu Model H10770). Mice were put on a restricted water regimen for the duration of the experiment [12]. During the task, mice were head-fixed but allowed to run freely on a circular treadmill, consisting of styrofoam ball suspended by air (PhenoSys, JetBall Virtual Reality System). Non-rigid image registration, ROI extraction, baseline de-trending, and spike deconvolution was performed offline. [13], [14]

C. Behavioral Task

The boundaries of two ensembles of two well isolated regions of interest (ROIs) were manually defined after a 5-15 minute baseline period of imaging. Depending on the density of viral labeling, we were able to observe 300-800 additional neurons in a 500µm² field of view. Ensemble activity was measured as the summed, running z-score normalized\( \Delta F/F_0 \) for each component neuron. This normalization is performed to scale each neuron by its dynamic range, in order to overcome potential differences in viral expression across cells, such that each direct unit would contribute equally to the effector. The cursor was smoothed with a running average of 2 - 3 frames. Closed-loop latency was 38ms, with a jitter of 17ms (95% confidence). Frequencies used for auditory feedback ranged from 1-20 kHz in quarter-octave increments to match rodent psychophysical discrimination thresholds [5]. Mice must modulate calcium dynamics in these neuronal ensembles to move the cursor to a high-pitched target tone that was associated with water reward, set at 2.5 standard deviations (SD) above the baseline (Fig 1b). Additional reward was prevented until the cursor returned to within baseline levels (1SD of the cursor mean).

D. L2-regularized Linear Regression Model

Indirect neuron features were derived from the spike deconvolved calcium activity within a 2 second sliding window [timepoints * spike and height features]. The response matrix consisted of the fluorescent change (\( \Delta F/F_0 \)) of each direct neuron (\( E_1a \), \( E_1b \), \( E_2a \), \( E_2b \)), the cursor (\( E_1 - E_2 \)), and the 2 ensembles \( E1 \) and \( E2 \). We separated the indirect neuron feature matrix (\( IND \)) and direct neuron response matrix (\( DN \)) into a training dataset and completely withheld validation dataset based at the moment of the ‘hit’. To avoid bias, we fixed the regularization coefficient- determined by cross-validating the regression procedure 5×. In each cross-validation iteration, 80% of the training dataset was used to estimate the model weights for each of 10 possible regularization coefficients of the remaining 20%. After the completion of cross-validation, a regularization-performance curve was obtained by averaging the cross-validation sample \( R^2 \) first across 10 regularization coefficients samples and then across all 7 direct neuron responses, then the best regularization parameter (21.544) was selected. Model estimated weights were used to predict responses of the withheld validation dataset. We repeated the modeling procedure using the features of indirect neurons to predict shuffled direct neurons’ response matrix in time (frames) axis. We shuffled the response matrix 1000× to build a null distribution of \( R^2 \) to obtain \( P \) values [15].

III. RESULTS

A. An Emergent Network Solution to Neuroprosthetic Tasks

Mice were able to modulate the calcium dynamics in chosen neuronal ensembles to move the cursor to a high-pitched target tone that was associated with water reward (Fig 1c) [5]. The task is designed to control for simplistic network level solutions by forcing two ensembles to be active differentially (i.e. \( E1 - E2 \)). However, an increase in reward frequency over a session could result from non-specific, or a multitude of degenerate network strategies- For example, upon hearing a tone, increased attention or arousal may increase excitability in individual cells, increasing network entropy compared to a baseline period when cells were less active. This feedback loop could increase the probability that

\[ \text{Dotted line} \quad \text{E} \quad \text{A} \quad \text{C} \quad \text{B} \quad \text{E} \quad \text{F} \quad \text{G} \quad \text{H} \quad \text{I} \quad \text{J} \quad \text{K} \quad \text{L} \quad \text{M} \quad \text{N} \quad \text{O} \quad \text{P} \quad \text{Q} \quad \text{R} \quad \text{S} \quad \text{T} \quad \text{U} \quad \text{V} \quad \text{W} \quad \text{X} \quad \text{Y} \quad \text{Z} \]
any network pattern would occur, compared to the baseline period. If this was true, both the variance across the network and the average number of events across individual cells in the interval immediately before the reward criteria is met should remain constant or increase over the course of the session, reflecting a strategy of using numerous state permutations to gain reward. However, we observe the opposite—total variance across the network tends to decrease over time, as does the mean population activity of indirect neurons in the moments leading up to the hit (Fig 2c-d). In contrast, the difference between the two ensembles that result in reward (i.e., $E1 - E2$) increases in magnitude over the course of a session on the same time interval (Fig 2a). Given these results, we next normalized the number of rewards over the session as a function of the number of rewards that would have been achieved if any other pairs of neurons had been driving the cursor. We find there is a strong trend to increase reward related network patterns over time, indicating that a limited number of specific activity patterns are being consolidated on this timescale (Fig 2f). Moreover, we find that the indirect neurons exhibit low dimensional structure that lead and follow rewarded patterns in direct neurons (Fig 2g), and this structure dramatically increases in consistency over the course of a session (Fig 2h-i).

B. Prediction of Output Cells Using Indirect Neuron Activity

Increase of the relative reward rate over time suggests that the network is learning to specifically modulate a subset of neurons to increase the reward rate (Fig 2g-i). We reasoned that it may be possible to observe network credit assignment by modeling how well the population is able to decode output neuron activity. Using L2-regularized linear regression (also known as ridge regression), we find that direct neurons $\Delta F/F_0$ can be accurately decoded from the indirect population activity with high accuracy ($R^2$ values for $E1_d = 0.77$, $E1_b = 0.78$, $E2_d = 0.71$, $E2_b = 0.53$) in addition to the ensemble activity ($R^2$ $E1 = 0.65$, $E2 = 0.69$). However, we find that our ability to decode direct neurons is not significantly different than the ability to decode any arbitrary neuron, or linear combination of neurons. In all cases, we find that a subset of neurons (10-20%) can decode the majority of the variance in the any given cells. However, the value of the cursor (simply $E1 - E2$) was significantly more difficult to decode compared to linear combinations of other neurons (cursor $R^2 = 0.36$), strongly suggesting that the network is not passively reflecting incoming sensory information. Interestingly, our ability to decode the cursor increases when considering the tone that will occur in the future ($R^2 = 0.7$ given $t + 400ms$), while over the same timeframe there is a decrease in decoding accuracy for other cells.

C. Spatiotemporal Distribution of Indirect Cells

Motor and premotor structures have been shown to be locally clustered when driving natural movement, which may result from intrinsic connectivity patterns [16], [17]. To test if this is true in our neuroprosthetic task, for every successful ensemble pattern that led to a reward, (ie. $E1 - E2 > 2.5SD$), we take the activity 1.5s before and 1.5s after this moment, and observe how many cells have consistent changes in fluorescence. We can compare the mean $\Delta F/F_0$ activity across a subset of ’hits’ (when direct neuron activity results in reward) for each indirect cell, and compared it to the mean ROI activity of an independent subset. (eg, even vs. odd hits). We find that roughly 10-20% of the population consistently forms a temporal relationship with the direct neurons (Pearson correlation is $> 0.9$), and consistently either leads or follows the activity of output neurons(Fig 4a,c) [10]. This result suggests that direct neurons are incorporated into large spatiotemporal sequences that satisfy the reward criteria. Plotting the spatial distribution of these cells reveals spatiotemporal clustering on the order of 100µm. (Fig 4b,d).

IV. DISCUSSION

Our findings recapitulate observations seen in across-day learning in both neuroprosthetic, and natural motor tasks:
behavioral performance correlates with an increased consistency of neural responses [8]. We observe an initial increase, then rapid decrease of unrewarded neural activity patterns in time. This result highlights a classic prediction from reinforcement learning theory: early in a session, there is high variance while the network explores various network patterns—those that lead to reward are consolidated and then exploited, and unrelated activity becomes sparse. In addition, we observe that the surrounding network can converge in space and time to produce sequences in cells that spatially cluster on the length-scale of 100\(\mu\)m. In this case, the coordinated activity of many neurons acts as the emergent functional unit of neuroprosthetic learning [18]. Possibly, this is the result of re-assigning pre-existing sequences that normally form motor primitives [19].

Interestingly, while the activity any arbitrary cell (or any linear combination, i.e., ensemble activity) can be decoded by the remaining population with high accuracy (mean \(R^2 = 0.70\)), prediction performance of the cursor is relatively poor (\(R^2 = 0.36\)). However, performance dramatically increases when considering the cursor value in the near future (\(R^2 = 0.66\) given \(t+400\)ms). This surprising result suggests that the network is not passively responding to a reward expectation based on ongoing sensory feedback, a point further supported by the observation that neurons display minimal modulation with passive tone playback [20]. Instead, this observation suggests that the local neural population contains both information about the current network state, along with an estimation of the upcoming sensory consequences resulting from that network state—a critical prediction of reward guided behavior.

**Figure 3.** Indirect neurons form robust spatiotemporal patterns leading up to, and following the rewarded output neuron configuration. (A) Three seconds of activity, centered at the reward criteria (‘trials’), were split into two interleaved groups (55 each, 110 total for this animal) and averaged across trials. The mean ROI activity was sorted on the even trials, and this sorting is applied to the odd trials. A subset of ROIs demonstrated high consistency (right panel displays 220/790 identified ROIs that all have \(>0.90\) Pearson correlation of their shuffled fluorescent time course). This correlated with peak amplitude (\(p<0.01\)). (B) Identified regions of interest colored by activity support songbird cortical sequences. (C) Neurons cluster in space and time with a length scale of 100\(\mu\)m, relative to the rewarded network pattern in direct neurons. (F) E1 cells are more highly tuned to the local population at the moment the reward criteria is achieved.

**References**


