Inferring the control signal driving zebrafish locomotion

A central objective in neuroscience is to understand how the brain orchestrates movement. Recent advances in automated tracking technologies have made it possible to generate rich behavioral datasets that can be exploited to gain insights into the neural control of movement. One approach to analyzing such data is to identify stereotypical motor primitives using cluster analysis. However, this categorical description can limit our ability to model the effect of more continuous control schemes. Here, we take a control theoretic approach to behavioral modeling and argue that movements can be understood as the output of a controlled dynamical system. Previously, models of movement dynamics, trained solely on behavioral data, have been effective in reproducing observed features of neural activity. These models addressed specific scenarios where animals were trained to execute particular movements upon receiving a prompt. In this study, we extend this approach to analyze the full natural locomotor repertoire of an animal: the zebrafish larva. Our findings demonstrate that this repertoire can be effectively generated through a sparse control signal driving a latent Recurrent Neural Network (RNN). Our model's learned latent space preserves features relevant to the fish's navigation while disentangling different categories of movements. Collectively the control signal and dynamics we identified offer a novel framework for understanding neural activity in relation to movement.

Behavioral dataset and Model Zebrafish larvae swim using discrete episodes of propulsion, known as swim bouts, which typically last around 200ms. We characterized the posture of the fish by measuring the bending angle along the tail (see Fig. 1A.) and compiled a dataset from observations of 100 freely-swimming larvae. Using methods from [1], we segmented the data into swim bouts which we classified into 11 distinct kinematic categories (see Fig. 1A.). To characterize zebrafish locomotion, we sought to learn dynamical systems that were able to generate its locomotor repertoire (see Fig. 1B.).

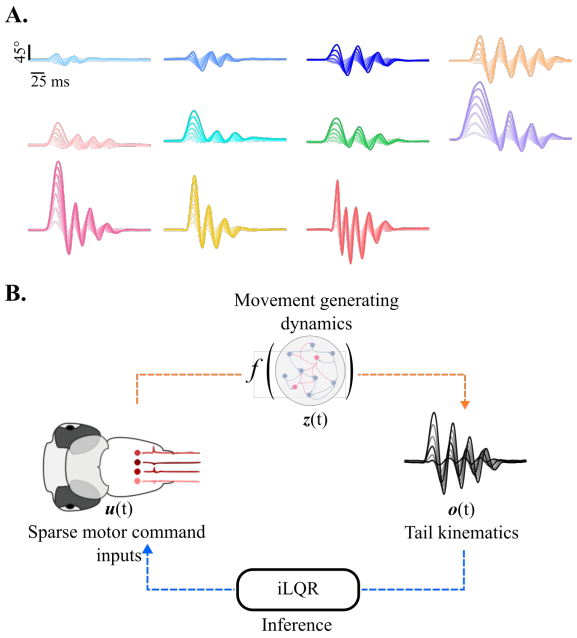

Fig. 1. (**A**) Samples of swim bouts from each category.(**B**) Illustration of the model setup.

Here, our goal is to identify both the latent control signals and the underlying dynamics that make up the complete locomotor repertoire of the zebrafish larva. Learning the dynamics of a system driven by unobserved inputs is a challenging system identification problem. We tackled it using the recently proposed iLQR-VAE method [2], to learn the latent RNN and infer the control signals from time series of behavioral observations. We assumed that the dynamics were driven by a *sparse* set of inputs. Specifically, we followed [2] and used a Student-t prior distribution. This choice is consistent with biological observations. Indeed, sparse electrical activation of brainstem neurons projecting to the spinal cord has been shown to be sufficient to elicit forward locomotion or escape [3,4]. The heavy-tailedness of the prior is thus well suited to inferring large sporadic inputs that may trigger swim bouts while encouraging the model to capture most behavioral segments via strong, near-automonous dynamics.

Low-dimensional, sparse control of larval zebrafish locomotion To find the models that best described the data, we varied the dimensions of the control signal and of the latent size, as well as the RNN architecture. We considered both linear and non-linear RNNs. We evaluated models based on both (i) their ability to reconstruct the data, and (ii) the sparsity of the inferred control signals. We found that all trained models could generate accurate reconstructions of our dataset ($R^2 > 0.94$). The high reconstruction fidelity can be attributed to the inherently low-dimensional nature of behavioral recordings. Yet, we found that matching the observations *using a very sparse control* was challenging, and possible only for the largest and most expressive models (non-linear RNN). Here, we found that the model's multivariate control impulse arose right before the onset of movement (see Fig. 2A.), thus

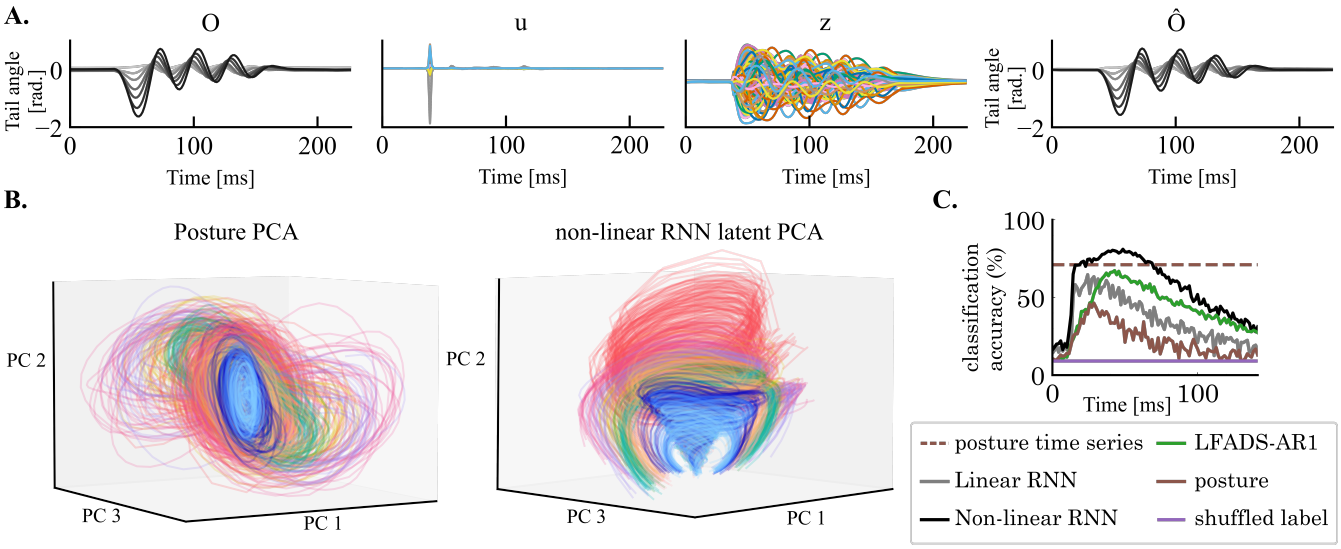

Fig. 2. (A) Example bout reconstruction ($R^2 = 0.97$ between the observation O and the reconstruction $\hat{O}$), with a sparse input u driving the latent trajectories z for the non-linear RNN. (B) PCA projections of the posture time series (left) and latent state (right) color-coded according to bout category. (C) Performance of linear classifier to predict the movement category using latent or posture data.

setting the initial conditions for the latent dynamical system. After this point, the bout was then generated by quasi-autonomous dynamics, before decaying back to a fixed point.

Most of the information to generate a movement was contained in this low-dimensional impulse. Indeed, restricting the input to its initial peak was still sufficient to reconstruct bouts with $R^2 = 0.82$ across the test dataset. Following this initial state, the movement unfolds by following the learned flow field. In the absence of an additional control signal, the trajectory in state space should therefore be untangled, with similar positions in state space leading to similar patterns in the near future [5]. We tested this in the models, by measuring how well we could decode the category of movement from a single snapshot of the latent state (Fig. 2C.). In contrast to postural trajectories, which were highly tangled, the low-tangling of the latent state space made it possible to classify movements accurately. Surprisingly, the state space of the RNN 40ms after the control peak provided a higher classification accuracy even when compared with a linear classifier trained on the full time series of tail movements (Fig. 2C.). It demonstrates the quality of the representation of the movement within the latent space.

To benchmark our method, we used LFADS [6] with an autoregressive prior for the control signal. The method successfully reconstructed the postural observation ($R^2 = 0.94$). However, we observed that in this regime, the latent dynamical system was predominantly input-driven. Indeed, we found a projection of the control signal displaying a correlation of $r = 0.73$ with the postural time series. This leakage from the input data resulted in a less informative dynamical system, as measured by the classification accuracy of the movement category from the latent trajectory(Fig. 2C.). This result suggests that for low-dimensional behavioral observation, sparse input priors are better suited to learn a meaningful dynamical system.

Control of spatial navigation The control signal accounts for the influence of sensory and decision-making areas. As such, this signal is expected to encode information relative to navigational landmarks such as the relative position of a prey that the fish want to capture. We propose that a straightforward mapping exists between the control impulse and the ensuing fish trajectory. This proposal is non-trivial, as our training dataset contained only posture information. Nonetheless, we found that spatial displacement can be linearly predicted from the initial latent state ($R^2 = 0.9$ for turn direction, $R^2 = 0.9$ for turn yaw and $R^2 = 0.72$ for swim distance). The control signal and dynamic inferred by our method therefore allow for a simple sensorimotor coupling.

References [1] Marques et al., *Current Biology* (2018) [2] Schimel et al., *ICLR* (2022) [3] Severi et al., *Neuron* (2014) [4] Xu et al., *Current Biology* (2021) [5] Russo et al., *Neuron* (2018) [6] Pandarinath et al., *Nature Methods* (2018)