

Background

Until this part of the semester, we've focused on how to think about other people's science through presentations and critique of primary literature. We've also practiced programming the behaviours of artificial agents in Webots using Python.

For this mini-project, we would like you to use these technical and conceptual skill by giving you hands-on experience using Python to analyse new data with the intention of discovering something new about how animals control their movements. Specifically, you will be exploring how signals between the brain and motor control centres control limb movements. We will be looking at two important classes of neurons:

Descending neurons (DNs) are defined by having their cell bodies and dendrites in the Brain and their axons primarily projecting to the insect motor control centre, the ventral nerve cord (VNC, roughly equivalent to the insect 'spinal cord').

Ascending neurons (ANs) are the opposite of DNs. They have their cell bodies and dendrites in the VNC and their axons project to the brain.

Previous theoretical studies in models and robots ([Ijspeert et al. Science 2007](#)) suggested that simply driving a small subset of DNs could flexibly elicit multiple behaviours (e.g., swimming and walking) by recruiting downstream motor circuits in new ways. Such a possibility could not be tested in real animals until recently when the building of transgenic strains of *Drosophila* (one strain in [Bidaye et al., Science 2014](#) and >100 strains in [Namiki et al. Elife 2018](#)) allowed scientists to target these neurons for studies of their anatomy and function and found that activation of individual DNs can trigger distinct behaviours, such as backward walking.

While exciting, strongly activating single neurons is a very artificial setting and these initial studies provided very limited details about how multiple DNs together controlled leg kinematics and behavioural dynamics as well as the role of ANs. Thanks to recent developments, it is now possible to record from many neurons connecting the brain and VNC at once using functional two-photon imaging ([Chen et al. Nature Communications 2018](#)).

Using this very recent technique, we recorded the activity of **123 neurons that connect the brain and the VNC at once**, while a fly is freely moving on an air-suspended spherical treadmill. You will be working with a novel dataset that no-one has ever analysed before.

With this data in hand, you have the opportunity to answer some fundamental questions regarding how groups/populations of neurons control behaviour in animals, and also to try out some new data analysis approaches to study the relationship between neuronal signals and behaviours in an unsupervised manner – taking “pesky” humans out of the loop in neuroscience.

The dataset

The dataset consists of recordings from one fly (R57C10-Gal4 > UAS-GCaMP6f, UAS-tdTom) across 12 trials of around 250s each. We provide the following two pandas dataframes which contain the data of all trials:

See the [Github repository for an interactive notebook to extract features of the data.](#)

The [actual data is on a Google drive](#) under the subfolder /Data.

Neural data: COBAR_neural.pkl

- Fluorescence traces of 123 neurons in the cervical connective expressing the genetically encoded Calcium indicator GCaMP6f ([Chen et al. Nature 2013](#)) imaged with a 2-photon microscope at a sampling rate of around 16 Hz. The exact time stamp of each sample is in the signal "t"

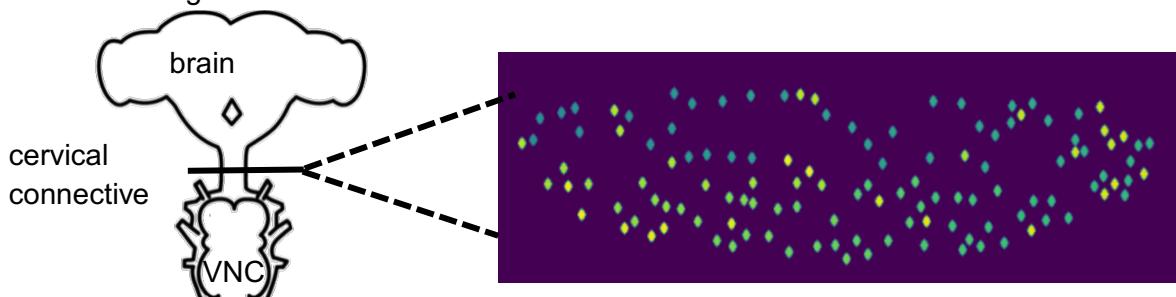


Figure 1: Left: The fly's central nervous system consisting of the brain and the ventral nerve cord (VNC). Right: Location of 123 neurons in a cross-section of the cervical connective.

Behavioural data: COBAR_behav.pkl

Behavioural variables are recorded/computed at a higher frame rate (100 Hz) to capture very fast movements of the fly. The behaviour of the fly is recorded with 7 cameras surrounding the fly. Using the marker-less pose-estimation algorithm DeepFly3D ([Günel et al. Elife 2019](#)), we track the joint positions over time. After alignment of the pose to a template fly (See [Lobato et al. BioRxiv 2021](#)), the joint angles are computed. Based on the joint angles, we trained a simple classifier, to predict the behaviour of the fly for every single frame.

- Joint positions of 5 joints per leg, each in xyz, including Coxa, Femur, Tibia, Tarsus, Claw → in total $5 \times 6 \times 3 = 90$ variables
- Joint angles. 7 per leg including Coxa_yaw, Coxa, Coxa_roll, Femur, Femur_roll, Tibia, Tarsus. → in total $7 \times 6 = 42$ variables
- Output of behavioural classifier. The classifier was trained on images of multiple flies including the following behavioural labels: abdominal grooming, antennal grooming, eye grooming, foreleg grooming, hindleg grooming, resting, walking. Resting and walking are the behaviours, which are observed the most. Sometimes, the fly pushes the ball with its abdomen.
 - o Prediction → the behavioural prediction of the classifier
 - o Entropy → A measure of uncertainty of the classifier. If high, then the classifier is not sure.
 - o Probability_Walking/... → the probability of each individual behaviour
- "twop_frame": This signal allows you to match behaviour frames (0:25199 per trial), to neural data frames (0:4039 per trial). The signal has a negative value if no neural frame matches the behavioural frame.

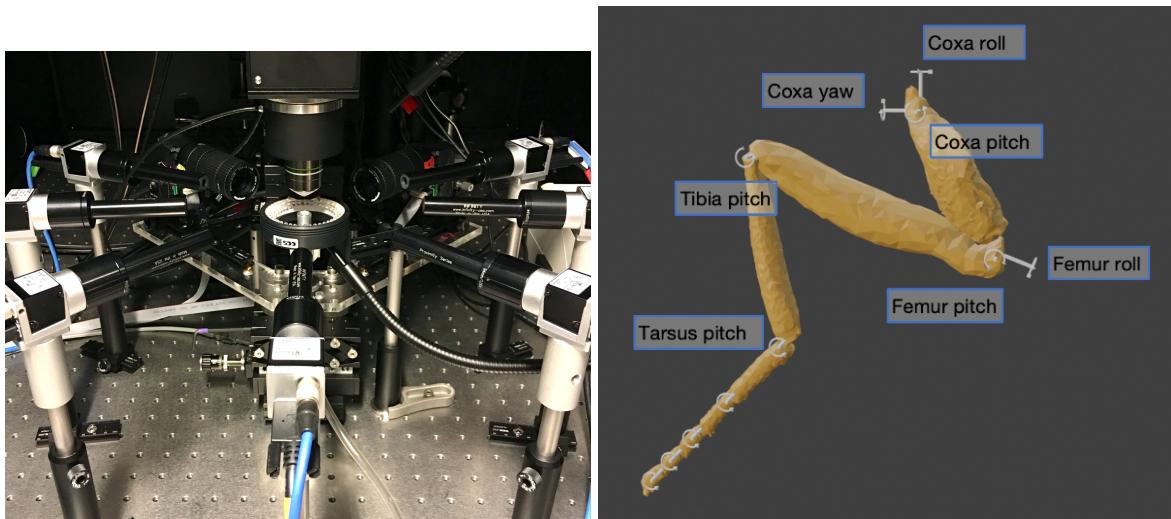


Figure 2: Left: recording set-up with 7 cameras surrounding the fly and the objective of the two-photon microscope
Right: A leg of a drosophila with the joint names and joint angles

Mini-project structure

Week 8 - 20.04.: Familiarisation with data, pre-processing, plotting

1. Familiarise yourselves with the data by plotting time traces
2. Data pre-processing
3. Understanding the types of behaviours, that a fly performs, while manually correcting behavioural labels of an automated classifier

Week 9 – 27.04.: Analysing behavioural/neuronal data by itself

1. Build a behavioural classifier
2. Perform dimensionality reduction & clustering on neuronal/behavioural data

Week 10+11 – 04.05.+11.05.: Correlating Neuronal activity with behavioural activity

Week 12 – 18.05.: Finalise data analysis & prepare mini-project report

Week 13 – 25.05.: Submit mini-project report & prepare presentations

Week 14 – 01.06.: Present mini-projects

We encourage you to be creative in analysing these data but at least address the following questions:

In general, we do not expect you to answer each individual question in the report/presentation. Make sure to think about each of the given questions and use them as a guide to design your data-processing / analysis pipeline. It should be clear from the report that you have spent some thoughts on each section. Rather than writing a question-answer style report, try to make the motivation of your analysis clear and put together a coherent story.

General questions for the report/presentation introduction

1. What is the experimental paradigm the data was generated with?
2. Why do you think the experiment were performed this way?
3. How might the experimental protocol, or data acquisition scheme have been better designed?
4. What are the advantages/disadvantages of studying spontaneous behaviour as opposed to a strict experimental design with many repetitions of the same experimental paradigm?

Week 8: Familiarisation with data, plotting

Part 1: Plotting the data

Look at the [following notebook on our Github for a guide to plotting](#)

1. Neural data

Plot time traces of individual neurons across multiple trials.

- a. Are they stable across all trials or do their activity patterns change?
- b. Are all neurons active during all trials or are some of them only active for some trials? What could be a reason for that?
- c. Are there some neurons which are much more active than others?
- d. Can you already find neurons which have very similar signals?
- e. What could be a good summary feature of a neuron? Its mean? Its maximum? Its standard deviation?
- f. Do all neurons have the same baseline fluorescence? If no, do you think this affects the analysis? Does the baseline change over time?

2. Behavioural data

Plot time traces of individual angles/joints across multiple trials

- a. Are they stable across all trials?
- b. Do some of them look very similar?
- c. Why do you think we want to look at joint angles in addition to the animal's pose?

OPTIONAL: Plot the 3d pose over time.

For your final report make sure that plots always have x- and y- axis labels and a title that is understandable. If you show multiple lines/data points in one plot, provide a legend and/or a colour bar. If you plot time series, always have time on the x-axis, not samples. Choose the scale of the x- and y-axis such that it is easy to see the message you want to convey with the plot.

Part 2: Data pre-processing

1. $\Delta F/F$

In the first part, you have observed different baseline fluorescence values. In order to account for that, one frequently calculates the so called $\Delta F/F$ ("Delta F over F"):

$\Delta F/F = (F - F_0)/F_0$, where F_0 is the baseline fluorescence.

- a. What would be a good way to compute the baseline fluorescence? Do you think taking the minimum provides a stable estimate? How could you stabilise this? (One idea might be to compute a moving average and then find the minimum or the 1% quantile).
- b. Should we use the same baseline for all trials or a different baseline for different trials?

2. Noise reduction

Measurements are inherently noisy. Filtering is a common strategy to remove noise. A couple of adequate methods might be low-pass filtering (e.g. Gaussian or Butterworth filter) or outlier removal (e.g. median filtering), but feel free to explore different methods. Use plotting as a means to see whether the signals are contaminated by noise and how your denoising affects them

- a. Do the signals look noisy to you? What could be sources of noise?
- b. How would you remove the noise? What do you consider noise, what signal?
- c. Does denoising limit your further analysis?
- d. How do parameters of the filters affect the signal appearance?
- e. In case you low-pass/high-pass filter the data, which cut-off frequencies did you use?

For your final report, provide examples that show your denoising approach and justify the choice of the parameters.

Part 2: Understanding different types of behaviour

To analyse behaviours, it is important to understand the types of behaviours the fly is capable of. We have previously trained an automatic classifier for the data, but it is not perfect. While familiarising yourself with the data, one of your tasks will be to annotate the behaviours of one trial (around 4 minutes) of behavioural recordings. **Each group will be assigned one trial** and can find the corresponding files [in the following google drive](#). We ask you to manually correct the automatic labels to:

- familiarise yourself with the behaviours that a fly performs when on a spherical treadmill.
- provide you with an even more exciting dataset for next weeks, where you can try to build an automatic classifier to classify behaviour and can get a better understanding of the relationship of individual neuron's activity with behaviour. Your colleagues will thank you for your labelling and you will be grateful for their work!

We have uploaded videos for each trial that include:

- the frame number,
- the time of the frame,
- a prediction from a not so optimal automatic classifier,
- the skeleton of the fly obtained by pose estimation superimposed on the video.

We suggest the following procedure:

1. Split up the trial amongst the two/three of you (i.e. person 1 does the first third and so on...)
2. Open the [Google sheets file](#) we provide. Go to the tab with the trial corresponding to your group. You should see something like this:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Date	Genotype	Fly	Trial	Frame	Manual	Prediction	Entropy	Probability a	Probability a	Probability e	Probability fc	Probability h	Probability rt	Probability walking	
2	210301	J1xC19		1	0	0	resting	0.04588146	0.00018622	6.77E-05	6.31E-05	5.00E-05	0.00076335	0.99301667	0.00585303	
3	210301	J1xC19		1	0	1	resting	0.03788282	0.00010744	4.22E-05	4.50E-05	4.33E-05	0.00064927	0.99440149	0.00471129	
4	210301	J1xC19		1	0	2	resting	0.04215738	9.77E-05	5.89E-05	4.14E-05	4.23E-05	0.00069808	0.99361343	0.0054482	
5	210301	J1xC19		1	0	3	resting	0.08693023	0.00015473	8.11E-05	7.60E-05	6.32E-05	0.0014947	0.98454088	0.0135894	
6	210301	J1xC19		1	0	4	resting	0.14975949	0.00086742	0.00018146	0.00015372	8.15E-05	0.00512976	0.97133694	0.02224923	
7	210301	J1xC19		1	0	5	resting	0.02738627	0.00025493	6.12E-05	6.59E-05	2.97E-05	0.00138638	0.99645751	0.00174443	
8	210301	J1xC19		1	0	6	resting	0.15533683	0.00084799	0.00015796	0.0001748	7.27E-05	0.02399906	0.96975515	0.00499238	
9	210301	J1xC19		1	0	7	resting	0.53454091	0.00104967	0.00052602	0.00058225	0.00016907	0.05300987	0.84974626	0.09491687	
10	210301	J1xC19		1	0	8	resting	0.72716019	0.00116573	0.00093445	0.00087925	0.00030854	0.08645749	0.7605777	0.14967685	
11	210301	J1xC19		1	0	9	resting	0.98798299	0.0028817	0.00152032	0.00113702	0.00049234	0.16928835	0.59253678	0.23214349	
12	210301	J1xC19		1	0	10	resting	0.70457788	0.00250639	0.0015661	0.00122597	0.00041019	0.02885294	0.73395579	0.23148262	
13	210301	J1xC19		1	0	11	resting	0.4687878	0.00367804	0.00237686	0.00193504	0.0007531	0.01342941	0.87138121	0.10644635	
14	210301	J1xC19		1	0	12	resting	0.13407128	0.00120521	0.00078812	0.00034034	0.00018006	0.00455807	0.97630982	0.01661837	
15	210301	J1xC19		1	0	13	resting	0.082866	0.00021749	0.00018557	0.00010726	4.85E-05	0.00149387	0.98566809	0.01227919	
16	210301	J1xC19		1	0	14	resting	0.06465497	0.00016527	0.00021894	8.73E-05	4.02E-05	0.00097812	0.98939496	0.00911528	
17	210301	J1xC19		1	0	15	resting	0.02980164	0.00010485	7.83E-05	4.29E-05	1.95E-05	0.00010803	0.99576909	0.00387738	
18	210301	J1xC19		1	0	16	resting	0.0683422	0.00020942	0.00011544	6.32E-05	2.98E-05	0.00017492	0.98797119	0.01143601	

The columns E, F, and G are the most important.

E contains the frame number, which you can also see printed in the video

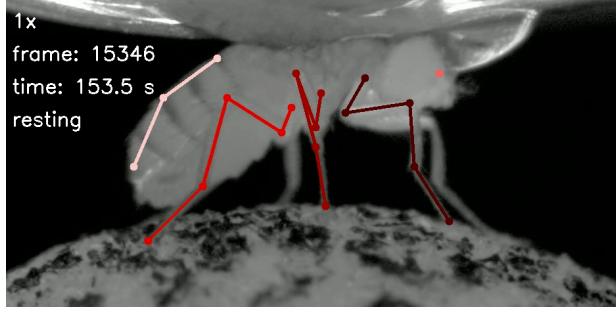
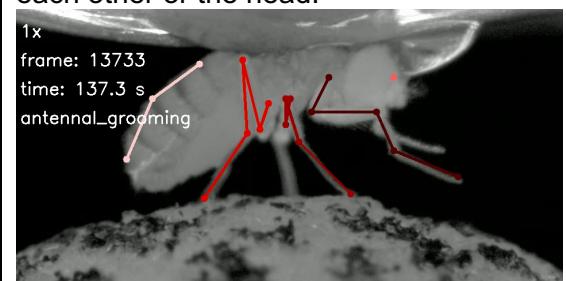
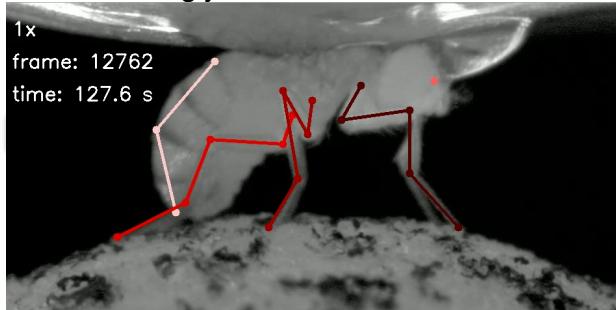
F is currently empty. This is where you should fill in the behaviours you find

G contains the automatically generated predictions (the following columns contain some more info about the classifier)

3. Open the video corresponding to your trial. We suggest to open it with a video player that allows you to move forward a single frame at a time. In Quicktime, this is possible by clicking the → and ← buttons on your keyboard. With VLC media player, [look at this website for help](#). Usually the 'E' key is setup as a so called hotkey to forward by one frame.

Please annotate the following behaviours:

- **resting**: when the fly is standing still on the ball without walking/grooming/pushing/...
- **walking**: when the fly is walking (irrespective of the direction) or turning
- **abdominal_pushing**: when the fly is pushing its abdomen against the ball (this behaviour is not included in the automatically generated labels, but is nonetheless frequent)
- **anterior_grooming**: when the fly is grooming (i.e. rubbing its front legs against) its eyes, antennae or front legs (this is supposed to summarise the automatically generated labels antennal_grooming, eye_grooming, and foreleg_grooming, because they are often hard to discriminate for the untrained eye)
- **posterior_grooming**: when the fly is grooming (i.e. rubbing its hind legs against) the abdomen or hind legs (this is supposed to summarise the automatically generated labels abdominal_grooming, and hindleg_grooming, because they are often hard to discriminate for the untrained eye). This behaviour is very rare, so don't worry if you don't find any instances.

<p>Example for walking: the fly is regularly, and in a coordinated manner, lifting multiple of its legs. Here: left front leg and right mid leg are lifted. Best judged by looking at the frames pre- and post-ceeding the current frame.</p>  <p>1x frame: 4997 time: 50.0 s walking</p>	<p>Example for resting: all legs are on the ball; no leg is moving (ball moves only slightly)</p>  <p>1x frame: 15346 time: 153.5 s resting</p>
<p>Example for anterior_grooming: both front legs are lifted up. They touch each other or the head.</p>  <p>1x frame: 13733 time: 137.3 s antennal-grooming</p>	<p>Example for abdominal_pushing: abdomen is bent downwards and pushed against the ball. Often, the other legs are move seemingly uncoordinated</p>  <p>1x frame: 12762 time: 127.6 s</p>

Tricks to speed up annotation:

- use copy and paste a lot, e.g., copy from the “Prediction” column, if the classifier was correct or drag down from the previous frame if the behaviour persists.
- Scroll through the video to identify changes in behaviour. Use the same annotation between two consecutive changes

While annotating, think about the following questions:

- What is “a behaviour”? Is it hard to differentiate between different behaviours?
- Are the behaviours you observe very stereotyped or do they vary along a spectrum?
- Would you have chosen different behavioural categories?

Potential questions for Weeks 9-11

(This is just an outlook to give you some ideas of what lies ahead. We will specify them before next week.)

Behaviour

1. How can you classify behaviour from 3D pose and joint angle time-series?
2. How can we define different “behaviours”? What timescales did you find were best for defining a behaviour?
3. Limb movements are very dynamic and a single time point will likely not capture all the necessary information about the current time point. Which methods could you use to extract features about the dynamics of the movement?

Neural Activity

1. Can you find neurons that have the same/similar signals?
 - a. How many are there?
2. What could be a reason for an organism to have multiple neurons conveying the same/similar signal?
3. Can you find “clusters” of neurons that have the same signals?
4. How many signal components do you need to describe most (e.g. 90%) of the variance in neuronal activity?

Correlating Neural activity with behaviour

1. What algorithmic strategies did you use to identify correlations between neural time-series and behavioural time-series?
2. What were the advantages and disadvantages of this approach?
3. What are alternative approaches you could take to overcome disadvantages of the approach you took?
4. Did individual neurons, or clusters of neurons relate to behavioural categories?
 - a. Can you classify behaviour from neuronal activity?
 - b. How many neurons do you need? Is a single neuron enough to reliably classify one or multiple behaviours?
5. If clusters in (4), how many clusters did you find? What clustering algorithm did you use?
6. Which behaviours were most represented by neurons? Which were not well represented (i.e., few or no neurons)?
7. For (6), why do you think you observed these results? What is the meaning in the context of how the brain is organized and what it is designed to do?
8. Did you observe neurons that were active during resting? If so, what do you think they might be used for?
9. Do you think descending and ascending neurons encoding behaviours serve different functions? If so, what are those differences?
10. Did you observe neurons that tracked the movements of individual joints (not larger scale behaviours)? If so, which angles? If not, what might be some technical limitations that would make this difficult to observe from this dataset?
11. What are some improvements one might make on the data quality or technical approach to increase the resolution of what we can conclude?
12. If you were given one year to analyse this and similar/more extensive datasets, what would you try to study?