UNIVERSITY OF LEEDS

School of Computing

FACULTY OF ENGINEERING AND
PHYSICAL SCIENCES

Final Report

Identifying neurons responsible for heat avoidance behaviour in C. Elegans

Julia Wilk

Submitted in accordance with the requirements for the degree of
BSc Computer Science

2023/2024

COMP3931 Individual Project

Summary

This project intends to identify neurons in C. Elegans worm which are responsible for its noxious heat avoidance behaviour. There have been prior attempts at this problem. Common methods to identify neurons that are a part of the heat avoidance circuit are lab-based and involve inhibiting a function of the tested neurons and checking whether the behaviour of the animal changes. As we don't have access to a lab, in this project we instead used statistical methods on a publicly available dataset in which half of the worms were exposed to a sudden noxious heat stimuli in order to identify neurons in which there has been a statistically significant difference in activity after the stimuli has been introduced.

We used the two-sample Kolmogorov-Smirnov test to compare distributions of averages of neural activity for each individual neuron in a small window after the stimuli between datasets of worms exposed to the noxious stimuli and datasets of control group worms. We eliminated high-variance neurons by only considering neurons which show a statistically significant difference in a small window after the stimulus is introduced and don't show a statistically significant difference in 99% of the time windows before the stimulus. The main achievement is identifying 17 neurons which pass these criteria for statistical significance. 5 of these neurons have previously been identified as being responsible for heat avoidance.

Acknowledgements

I'd like to thank my supervisor, Professor Netta Cohen, for her support and advice with the project. Our conversations left me with plenty of interesting ideas to pursue further after the end of this final year project, and her introduction to C. Elegans gave me an appreciation of the research on the fascinating world of neuroscience.

Table of Contents

Chapter 1
Introduction and Background Research

1.1 The choice of C. Elegans

Caenorhabditis Elegans is a small non-pathogenic roundworm that lives in the soil and feeds on microbes (1). Despite being seemingly insignificant to humans, there have been thousands of papers written on it (2). The first reason for the prevalence of C. Elegans in research is that it provides a great model to study the nervous system. It's one of the simplest organisms with a nervous system, but despite its simplicity, it exhibits many of the biological functions that are central problems of human biology: learning, aging, nerve function, and avoidance and attraction behaviour (1)(3). This results in a good trade-off between simplicity and complexity: understanding processes that take place in the brain of C. Elegans is more feasible than with more complex animals, and it can shed some light on how these processes are carried out in the latter. The seconds reason for the popularity of C. Elegans is that it's a good lab animal due to its short life cycle, transparent body, constant cell number and position between individuals, and the ability to grow on Petri dishes (4).

C. Elegans has a "brain" which consists of 302 neurons connected via ~7000 synapses (5). Unlike humans whose neurons use sodium and potassium ions to encode their activity, C. Elegans neurons mainly use calcium ions (6). The concentration of calcium ions tends to rise and fall gradually, and most neurons in C. Elegans don't produce "spikes" (7)(8).

Many of the C. Elegans neurons come in pairs, with one neuron from each pair located on the left and right side of the body (denoted by "L" or "R" at the end of the neuron name). The neurons in a pair belong to the same neuron type and have similarities in connectivity and composition, but they can respond to stimuli in different ways. (9)(10)

1.1.1 Mysteries to solve in C. Elegans

There is a lot of information currently known about C. Elegans, such as all the neurons and the synaptic connections between them (11), and the type and function of many of the neurons (9). Using this knowledge and brain recordings as well as behavioural responses from experiments, we can speculate on "circuits", which are sets of neurons responsible for certain behaviour and the mechanism in which they produce this behaviour. As an example, an olfactory avoidance circuit has been identified in C. Elegans, which consists of sensory neurons detecting olfactory stimuli, interneurons which process this data to determine

whether the sensation is strong enough to trigger an escape response, and motor neurons which send signals to muscles responsible for backwards movement (54).

However, there are still many unsolved mysteries in C. Elegans. Even though the synaptic connections between neurons are known, the strength of these connections has not been established (12) and there are many neurons which are not directly connected by a synapse but they can stimulate each other through extrasynaptic signalling (13). Furthermore, a large number of neurons change how they encode behaviour over time depending on the state they're in (14). Therefore, even though neural mechanisms and sets of neurons responsible for some behaviours have been identified, it doesn't provide an exhaustive explanation of how the neural system of C. Elegans produces behaviours.

1.1.2 Temperature avoidance in C. Elegans

C. Elegans are an ectotherm, which means that they can't regulate their own body temperature and rely on the temperature of the environment. Because of this, they have a preferred temperature. (15) Given a gradient of temperatures, they will move towards the temperature they grew up in (16), and neural mechanisms responsible for this behaviour have been found (17).

However, a response to a sudden significant rise in temperature is different from this mechanism, because it triggers an avoidance response – the worm starts escaping. There's an interest in studying the thermal avoidance mechanism in C. Elegans in order to further pain research in vertebrates, because the worms' responses to noxious heat can be pharmacologically manipulated by similar chemicals to those used in pain research, e.g. capsaicin (53).

There has been research into the heat avoidance behaviour in C. Elegans, and neurons that are likely to be a part of the heat avoidance circuit have been identified along with the mechanism that gives rise to the avoidance response (18). This mechanism involves sensory neurons "AFD", "AWC", and "FLP" responsible for detecting thermal stimuli (19), the last of which releases a neurotransmitter to stimulate neurons "AVA", "AVD", and "AVE" which promote backwards locomotion (20). Furthermore, the following neurons have been identified to be linked to heat avoidance response: "ASJ", "AWB", "ASK", and "ASH" (21). Functional redundancy has been identified in the neurons associated with heat avoidance (21).

1.2 Methods of detecting significant neurons

The neurons in the heat avoidance circuit as described in the section above were identified by inhibiting the function of these neurons e.g. by microsurgery or by producing genetically mutated worms, and subsequently observing whether the worms' response to noxious heat stimuli was reduced (19)(20). Another method was used to help identify aggression-related neural circuits in fruit flies, where stimulating a specific neuron led to the fly being in an aggressive state for 10 minutes (18).

However, due to the lack of access to an experimental lab, we used other methods to identify significant neurons. We used publicly available datasets of neural signals in C. Elegans collected from 40 animals, about half of which were exposed to a noxious heat stimulus (see Section 1.3 for more detail). We refer to these two groups of datasets (the datasets of worms exposed to the stimulus and the datasets of control group worms) as "heat" and "baseline".

Statistical methods such as analysis of variance (ANOVA) have been used to identify changes in pattern of neural activity. For example, ANOVA has been used to detect event-related potentials (a change in pattern in human brain activity) from EEG recordings of a human brain after an introduction of stimuli. The method was used in research to identify EEG channels that had a statistically significant change in recorded activity in a post-stimulus time window. (22) In the context of the datasets of the neural activity of C. Elegans that we use, this method could be adapted to identify neurons which had a statistically significant change in activity.

However, ANOVA is limited in that it only compares the differences in means between two distributions, due to an assumption that variances are approximately equal (23). This means that ANOVA would not detect a statistically significant difference between two distributions with same mean but very different variance. Furthermore, ANOVA would not detect differences between distributions of different shapes, due to its assumption of the data being normally distributed (23).

1.2.1 The Kolmogorov-Smirnov test

A statistical test which considers the differences between general shapes of distributions, including the differences in means and variances, is the Kolmogorov-Smirnov test. The KS test calculates whether two sets of samples came from the same underlying distribution. (24) In the context of our datasets, the KS test could be used to compare whether the heat and baseline neural signal values in a time window after the stimulus are likely to have come

from the same distribution (which would imply that there is no statistically significant difference), or from different distributions (which would imply a statistically significant difference between them).

In order to compare two distributions, as part of the KS test each distribution is turned into an Empirical Cumulative Distribution Function (EDF). EDF is a function that for each value of x, the value of y represents the fraction of the values in the distribution that have a value equal or less than the value of x. After turning each distribution into an EDF, the KS test calculates the value of the maximum distance between EDFs of two distributions. Because we compare two distributions of gathered data (rather than comparing one distribution of gathered data to a reference distribution), we use the two-sample version of the KS test. (24)(25)

The KS test returns two values: D-statistic and a corresponding p-value. The D-statistic (where “D” stands for “distance”) represents the maximum distance between EDFs of two distributions, and the p-value represents the probability that the result arose due to random chance. To calculate the p-value, D-value and size of the distribution are used. Therefore, only considering the p-value is enough to determine whether there is a statistically significant difference between the two distributions. (26)(27) The accepted threshold for determining that a result is statistically significant is p-value less than 0.05 (28).

The two-sided Kolmogorov-Smirnov test has been used in research by Randi, F., Sharma, A.K., Dvali, S. et. al. to investigate which pairs of neurons in C. Elegans are functionally connected. The KS test was used to compare the time series data of neural recordings from a time window after a stimulus (in this case the stimulus was another neuron being activated) to recordings where no stimulus was introduced. (13) This is similar to our intention of comparing heat and baseline time series data of neural recordings.

The KS test has the assumption of individual data points not being correlated. However, that is not true for any time series data, as value at time T will be in some way related to the value at time T-1 (a phenomenon called “lag-1 autocorrelation”). (30)(31) In the abovementioned research by Randi, F., Sharma, A.K., Dvali, S. et. al. (13), the authors address this by taking an average over a post-stimulus time window for each recording, which turns each set of individual recordings in a time window into one value. Then, the two distributions which are compared by the KS test are the distribution of averages of neural recordings where the worms were exposed to the stimulus, and the distribution of averages of neural recordings where the worms were not exposed to the stimulus.

1.3 The datasets

We used publicly available datasets downloaded from wormwideweb.org. The datasets were published in 2023 as part of C. Elegans brain-wide representations of behaviour research by Atanas et al. The datasets contain recordings of neurons of 40 worms over 16 minutes, with each dataset corresponding to one worm and 100 – 150 neurons. Most of the neurons were labelled so that the name of the neuron is known, however that was not possible for some neurons in each of the datasets. Furthermore, some of the labelled neurons contain “?” in the name, indicating that the type of the neuron is known, but it’s unknown whether it’s a left or right neuron. About one half of the worms were exposed to a noxious heat stimulus at minute 8, and the remaining worms were the baseline control group which was not exposed to that stimulus. The stimulus was delivered by a laser which heated up a worm’s environment from room temperature to over 50℃, which returned to baseline temperature after 3 seconds. The datasets contain z-scored time-series data of calcium levels in each recorded neuron, which represent the level of activity in that neuron over the span of 16 minutes. (14)

Chapter 2
Methods

2.1 Preparing the data

The time series datasets were all recorded over the span of 16 minutes, and they all had around 1600 individual measurements. However, many datasets had a number of measurements slightly lower or higher than 1600. This is not because they were recorded over a shorter or longer timespan, but rather because the time between individual measurements was slightly different for each dataset.

In order to calculate averages later on, we had to make all datasets have the same number of datapoints. To do that, we created an empty list of 1600 points for each dataset (which, after populating it, will form the new resized dataset), and then selected the value in the original dataset that was measured the closest in time to each of the 1600 points in the new resized dataset. The following is a further explanation of the method with an example. 1600 datapoints assumes that datapoints were measured 0.01 minutes apart. Therefore, for example, if we're looking to populate the 120^{th} datapoint in the new resized dataset which corresponds to T = 120 * 0.01 = 1.2 minutes = 72 seconds, and in the original dataset there are individual measurements taken at T_A = 70 seconds and T_B = 75 seconds, then we will select the neural signal value measured at T_A to populate the 120^{th} datapoint in the new resized dataset, because T_A is closer to T that T_B is.

Neurons without a label and neurons with "?" in the label are not considered. The latter is because there can be a significant difference in the signals between left and right neurons of the same type (10). Furthermore, only neurons with at least 5 measurements within each group are considered in order to reduce the impact of outliers.

2.2 Calculating differences within group versus between groups

For each individual neuron, there are two groups of datasets: heat and baseline. If there is a large difference in values between heat and baseline, especially around the time when the heat stimulus was introduced, this can imply that this neuron is significant for heat avoidance response. However, some neurons might always exhibit a lot of variability in their values, both between two groups and within each group, and therefore a large difference between heat and baseline for these neurons might not mean much. This is why we calculated the difference within each group as well.

2.2.3 Calculating the average

To calculate the difference between groups, we first calculated the average within a group for each neuron as follows: for each of the 1600 points, we calculated the average neural signal value for that point, across all heat datasets and all baseline datasets separately.

2.2.4 Differences between groups

After calculating the average, we considered the differences between groups in three time windows: across the whole 16 minutes, between 6 and 10 minutes (which includes the stimulus time at 8 minutes), and from 8 minutes onwards. For each time window, we took the absolute difference between heat and baseline average value for each of the pairs of corresponding datapoints, then sum these differences across all pairs of datapoints considered (1600 for the first window, 400 for the second, and 800 for the third). Lastly, we normalised the result by dividing the sum by the number of pairs of datapoints considered, which results in an average absolute difference per datapoint for each time window. This is so that we can compare the differences across the different time windows.

2.2.5 Differences within groups

For each neuron, we considered datasets in each group separately. We considered all possible unique pairs of datasets within each group, and calculated the absolute difference between heat and baseline for each of the 1600 pairs of corresponding datapoints. These were summed across all unique pairs within a group, then divided by the number of unique pairs to give an average absolute difference within a group. This was normalised by dividing it by the number of datapoints considered, that is 1600.

2.2 Visualisation

We used matplotlib Python module to display the averages of heat and baseline datasets for each neuron, and to display all the individual measurements that were used to calculate the averages. We ordered the neurons by the largest difference between groups in the time window 6 – 10 minutes. This is to show neurons that are likely to be significant in their response to the heat stimulus, because the time window 6 – 10 minutes includes the time when the stimulus was introduced. We displayed the top 10 neurons.

The purpose of the visualisation is to check for clear changes in the pattern of neural signals, and to cross-validate and put into context the results from statistical methods. Visualisation in itself does not constitute proof of a neuron's significance.

It's important to note that not all changes in pattern would be visible in taking an average within a group. As an example, the following figure demonstrates a change in pattern that would be "cancelled out" in an average:

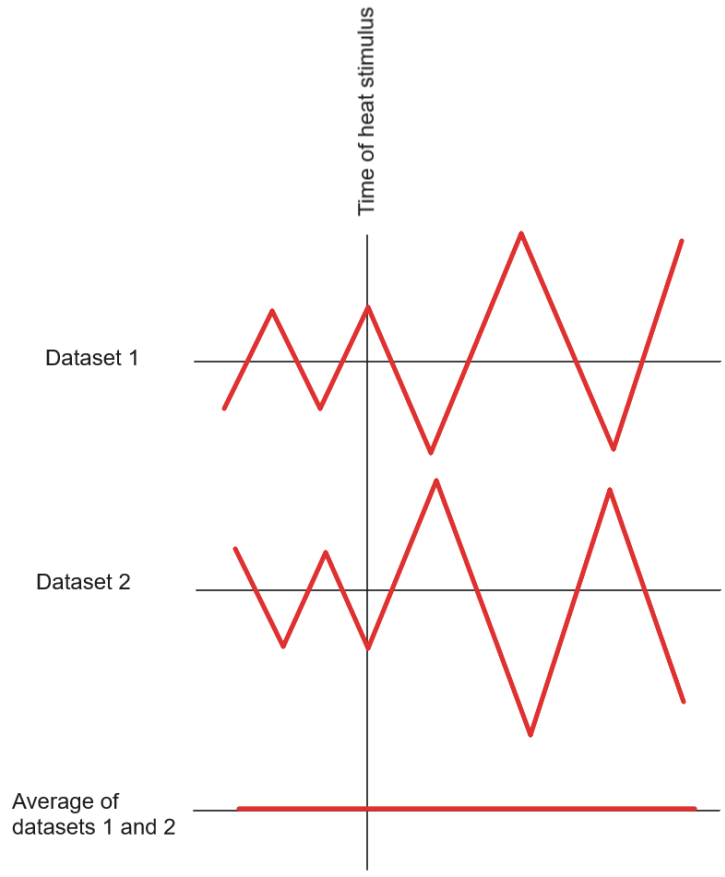


Fig.1: An example where a change in pattern is not reflected in the average. Dataset 1 and dataset 2 contain recordings of neural signals which have the same absolute values, but one value is negative when the other is positive which leads to them being "cancelled out" in the average. Even though the pattern clearly changes after the heat stimulus (through the strengths of neural signals increasing to about twice their size in both datasets), the change in pattern is not reflected in the average.

Therefore, as helpful as averages are to visualise and contextualise the data as well as notice potential changes in pattern, other methods are necessary to identify statistically significant neurons.

2.3 Statistical methods

2.3.1 Finding a statistically significant difference after the stimulus

To find neurons that show a statistically significant change in signal after the stimulus is introduced, we used a two-sided Kolmogorov-Smirnov test. To carry out the KS test, we used the “ks_2samp” function from the scipy.stats library in Python.

The test compares distributions of heat versus baseline neural signals for each neuron, where each distribution consists of averages of neural signals in a post-stimulus time window for each individual dataset within the group. Four time windows within which the average is calculated were considered: 1s, 5s, 10s, and 20s. Different start times for the time window were considered between 0 and 20 seconds from the introduction of the stimulus, with a time step between start times being the size of the window (hence for the window size of 20s only one start time was considered at T=8min, for the window size of 10s two start times were considered at T=8min and T=8min 10s, etc.).

The following figure illustrates the process while considering an example neuron “X”, for which there are 3 baseline datasets and 3 heat datasets (in practice, only neurons with at least 5 datasets in each group were considered). For simplicity, only window size of T = 5s is considered.

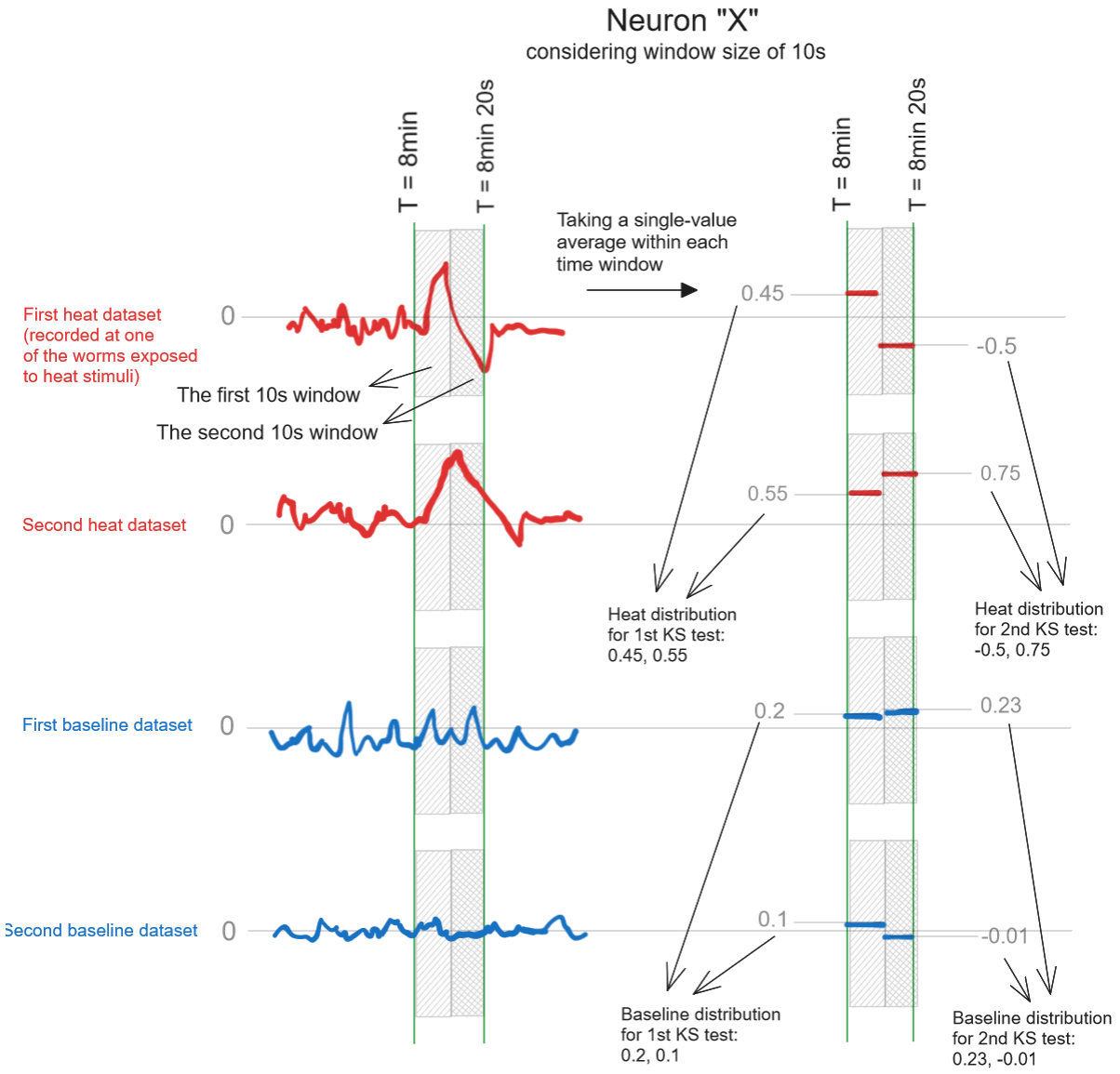


Fig. 2: Visualisation of performing a KS test, where the window size for calculating an average is 10 seconds and the timeframe of 20 seconds after the stimulus is considered.

For the neuron to be considered potentially significant, the KS test had to return a p-value of less than 0.05 for at least one of the window sizes (and for at least one of the start times within that window size).

If a neuron was potentially significant, the difference between the stimulus time and the start time of the window in which p-value reaches below 0.05 for the first time was recorded (called “delay”), as well as the time difference between the start times of the first and the last window in which p-value < 0.05 (called “duration”, with a maximum value of 2 minutes).

2.3.2 Eliminating high-variability neurons

If a neuron was considered potentially statistically significant under the criteria in Section 2.3.1, then an additional check was carried out to eliminate neurons which have high variability (and therefore, for these high-variability neurons, a KS test would likely return $p < 0.05$ for time windows before the stimulus as well).

The check was carried out as follows. For each of the statistically significant time window sizes identified in Section 2.3.1, we performed a KS test on windows of that same size in the timeframe between 0 and 8 minutes, with the time step between start times of each of the windows being equal to the window size (therefore for window size of e.g. 20 seconds, the KS test would be performed on windows 0-20s, 20-40s, 40-60s, …, 7min 20s – 7min 40s, 7min 40s – 8min). If 99% or more of the KS test results on these windows returned p-value higher than 0.05 (indicating a statistically insignificant result), then we considered that neuron to be statistically significant. In other words: neurons considered to be statistically significant were those neurons that did not have more than 1% statistically significant values for time windows before the stimulus is introduced, but did have a statistically significant value for some of the windows in the time frame between 0 and 20 seconds after the stimulus is introduced.

2.4 Visualising p-value as a function of time

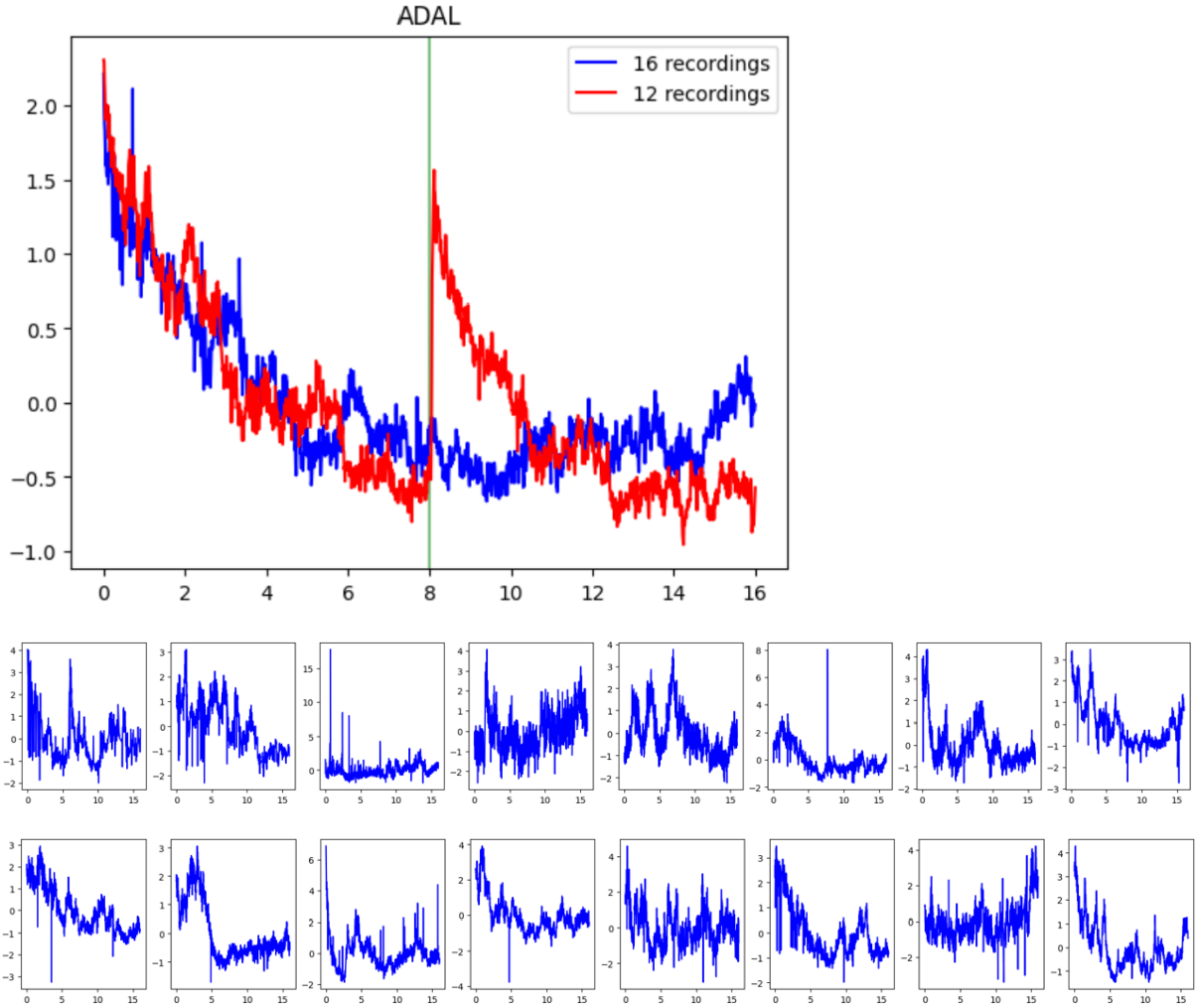
As part of Section 2.3, we calculated p-values for each window size over the whole 16 minutes (but only the p-values for windows in the timeframe 0s – 8min 20s were considered for identifying statistically significant neurons). For context and visual aid, we displayed a graph of p-values for windows calculated over the whole 16 minutes, for each window size that was considered statistically significant after passing all the criteria outlined in Section 2.3.

Chapter 3
Results

3.1 Visualisation

In 7 out of the top 10 visualised neurons, a clear change of pattern is visible. Different changes in pattern of neural activity can be observed for different neurons: a rapid increase which gradually goes back down, a rapid increase which quickly goes back down, a sustained decrease, a rapid increase followed by a sustained decrease back, and a rapid decrease followed by a quick increase followed by a gradual decrease back.

In 2 out of the top 10 visualised neurons (“RIMR” and “RIML”), there is no clear change in pattern. In 1 of the neurons (“AQR”) there is a clear change in pattern after the stimulus, but there is also a clear change in pattern shortly before the stimulus.



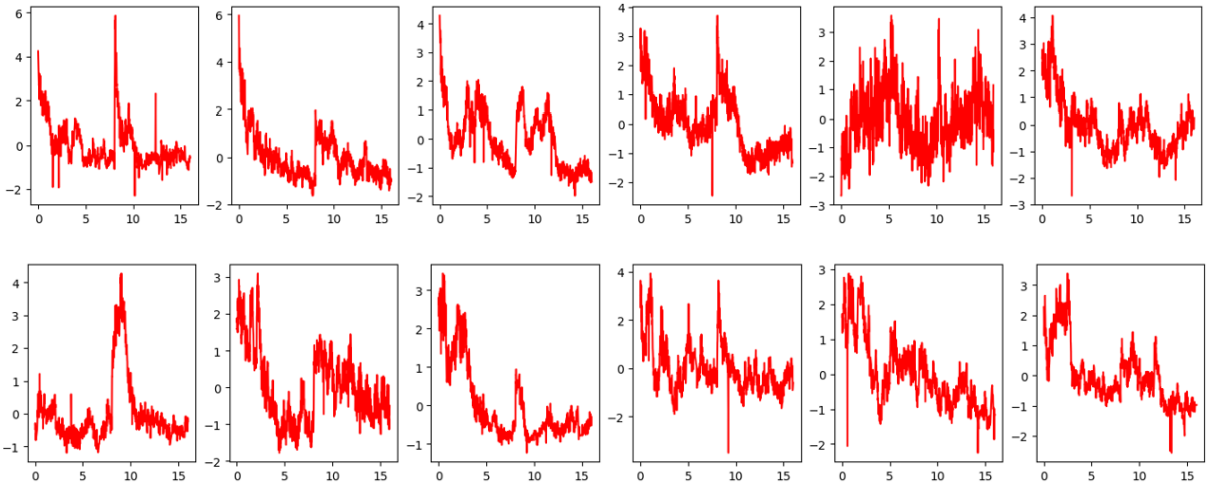
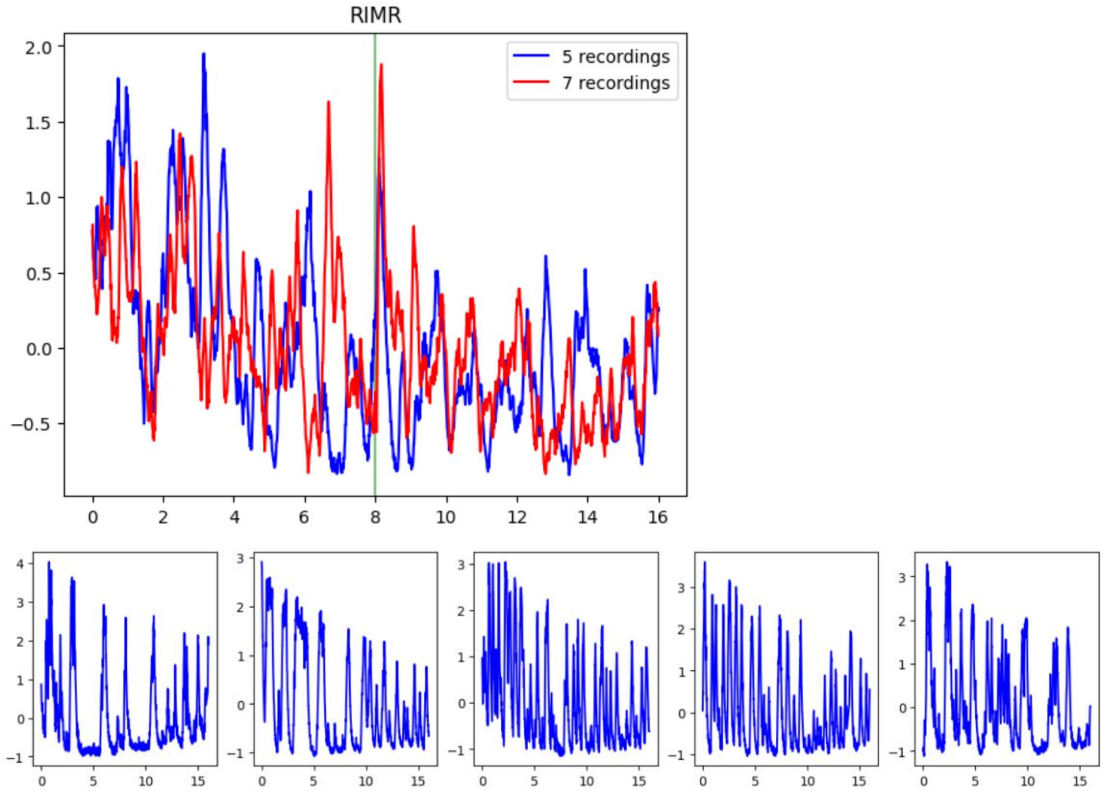


Fig.3: An example of a visualisation where a clear change in pattern after 8 minute mark (green vertical line in the top chart) is visible, for neuron “ADAL”. X-axis represents time (in minutes), and y-axis represents the strength of neural signal (in standard deviations). Top: The comparison of averages between heat and baseline datasets. The change of pattern is a rapid increase in activity which gradually goes back down. The average is calculated across 16 baseline recordings (blue) and 12 heat recordings (red). Two rows of blue figures: All individual baseline recordings used to calculate the average. Two bottom rows of red figures: All individual heat recordings used to calculate the average.



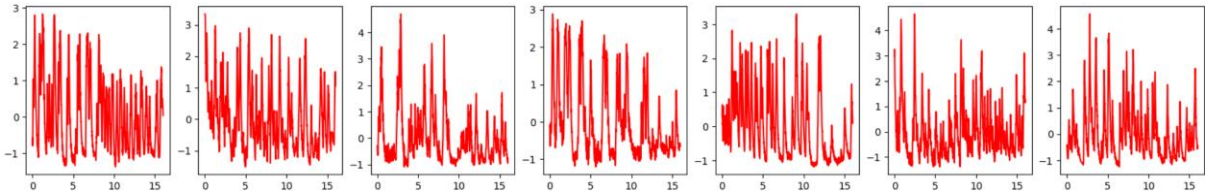
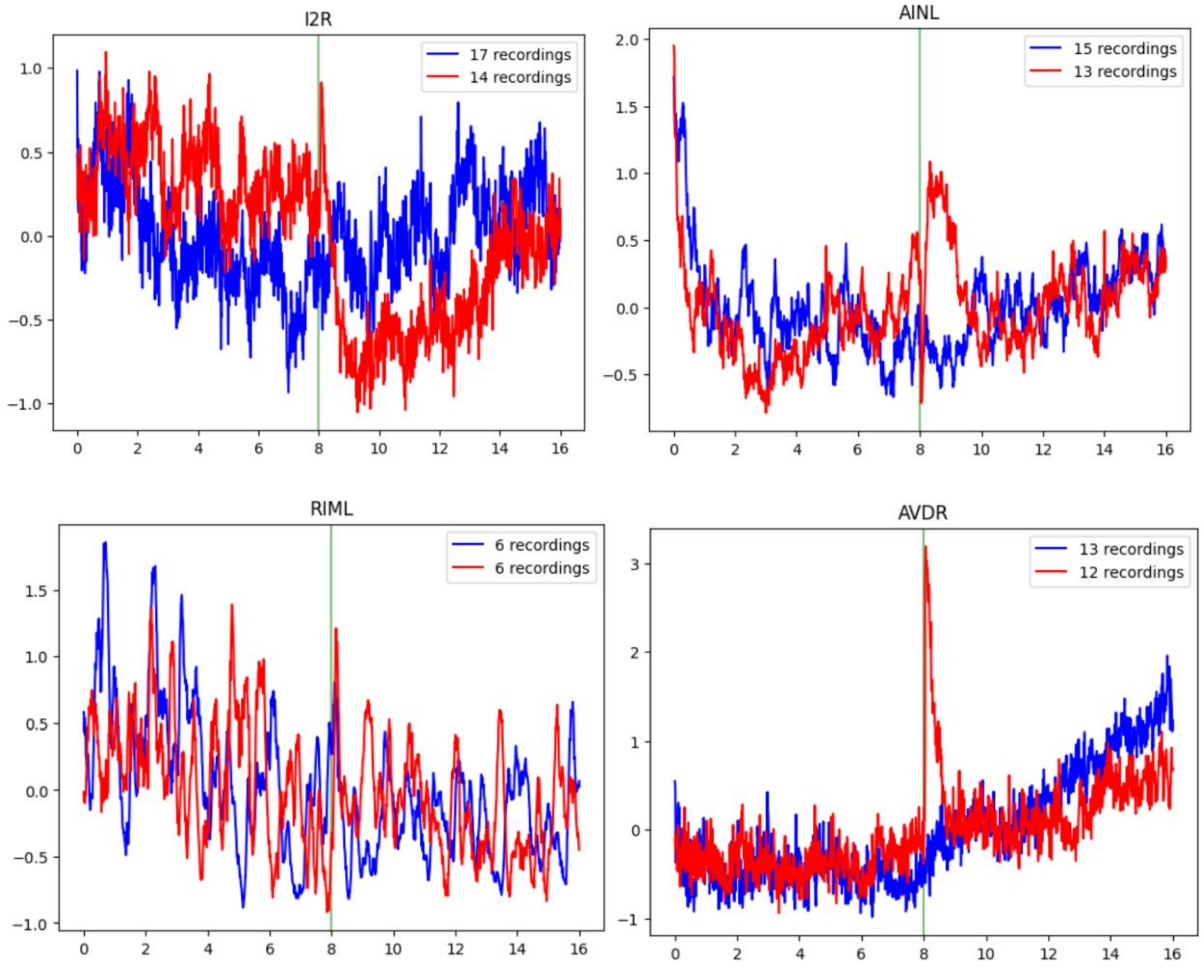


Fig.4: An example of a visualisation where no clear change in pattern at minute 8 (green vertical line on the top chart) is visible, for neuron “RIMR”. X-axis represents time (in minutes), and y-axis represents the strength of neural signal (in standard deviations). Top: A comparison of averages between the heat group (red) and baseline group (blue). Middle: All individual baseline measurements used to calculate the baseline average. Bottom: All individual heat measurements used to calculate the heat average.



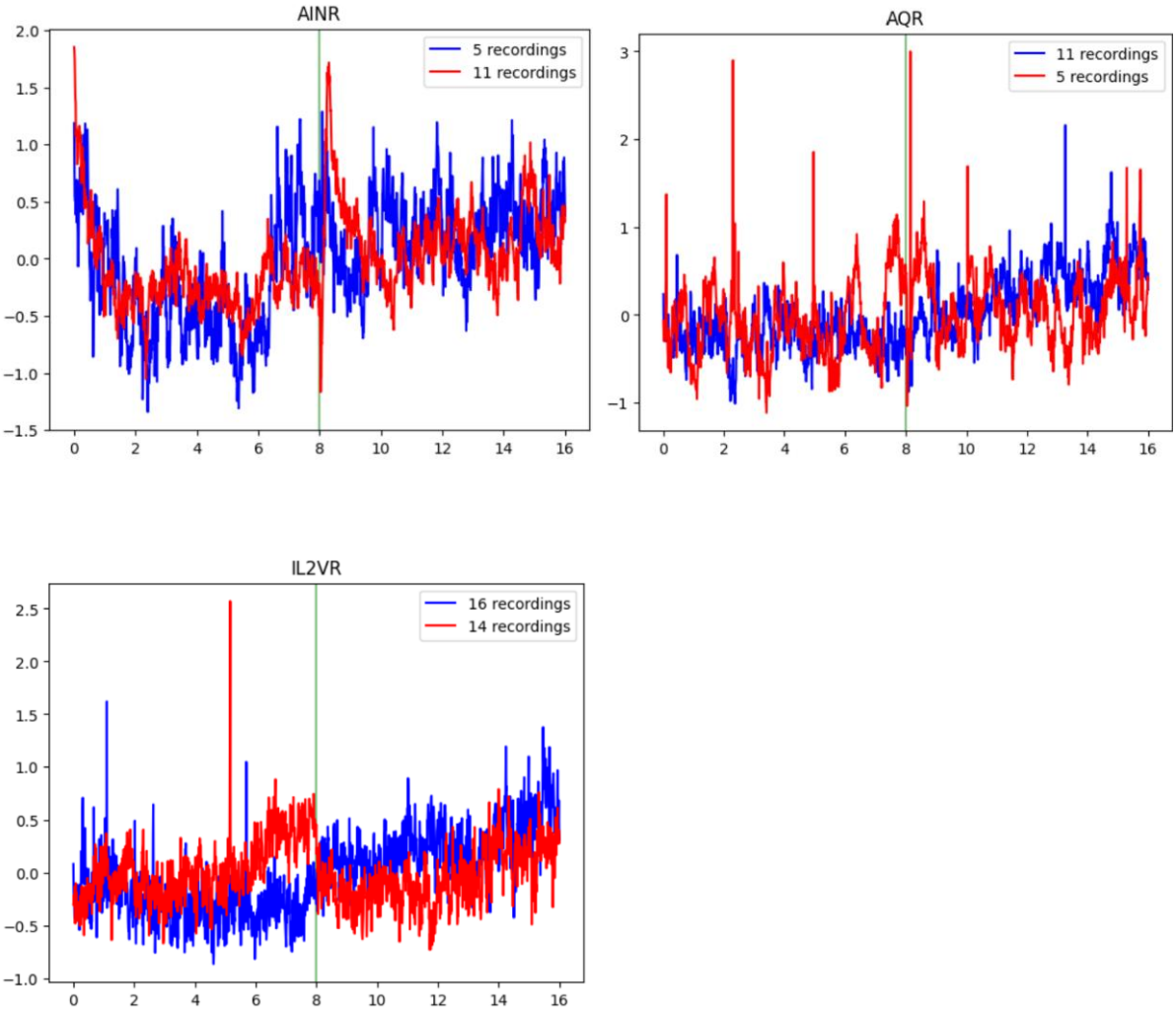


Fig.5: Visualisations of the remaining top 7 neurons. X-axis represent the time in minutes, y-axis represents the strength of the neural signal in the number of standard deviations, blue values represent the average of baseline datasets for a particular neuron, red values represent the average of heat datasets for a particular neuron, and the green vertical line represents the time of the noxious stimulus.

For visualisations of averages of top 10 neurons along with the individual measurements used to calculate the averages as well as the normalised values of the differences within and between groups, see Appendix C.

3.2 Calculating differences within group versus between groups

64 out of 109 of all considered neurons show the highest difference between group averages to be in the timeframe 6 – 10 minutes.

In general, the average differences within group are consistently larger (around two times as much) than the differences between group averages. The normalised differences within group have values close to 0.4 – 0.5, and the normalised differenced between group averages have values close to 1 for most neurons.

For the top 10 neurons with the highest difference between groups in the timeframe 6 – 10 minutes that don't show a clear change in pattern in visualisation which indicates high-variability neurons ("RIMR" and "RIML"), there is no significant increase in the normalised difference within groups, in comparison to the neurons with a clear change in pattern.

3.3 Statistical methods and p-value as a function of time

17 neurons were identified that passed all the criteria for statistical significance as outlined in Section 2.3. For most neurons there was only one time window that fulfilled the criteria.

Out of 13 neurons for which at least some functions are known, 5 have known associations with either backwards locomotion ("AVDL" and "AVER") (32)(33) or sensing noxious temperatures ("ASHL", "FLPR", "AWBL") (34)(21)(36). Of the remaining 8 neurons, one ("ADLL") has a known association with responding to general aversive stimuli (37)(38), and two ("ASGR" and "IL2R") have known associations with responding to harsh environmental conditions (39)(40). Of the remaining 5 neurons, four ("NSML", "MCL", "AUAL", and "RIVL") have associations with feeding behaviours (41)(42)(43)(44), and two ("RMGL" and "AUAL") have associations with social behaviours (45)(46)(42).

9 of the 17 neurons have sensory functions, 8 have interneuron functions, and 4 have motor functions. All sensory neurons with a connection to noxious stimuli show a delay of 0 seconds. Other sensory neurons show a delay of 0 – 20 seconds. Sensory neuron "IL2R" that regulates behaviours associated with surviving under harsh conditions (40) shows a delay of 20 seconds. Most interneurons, including interneurons associated with backwards movement, show a delay of 0 seconds. One of the interneurons ("RMGL", associated with social behaviours (45)(46)) shows a delay of 10 seconds. Motor neurons show a delay of 0 – 10 seconds. None of the motor neurons identified have associations with noxious stimuli or backward locomotion.

Durations vary between 5 and 120 seconds. Neurons that are certain to have a sensory function have durations 20 – 45 seconds, and neurons with a possible sensory function have durations 80 – 86 seconds. Interneurons have durations 25 – 120 seconds, with

interneurons responsible for backwards locomotion having durations between 40 and 105 seconds. Motor neurons have durations over 86 and up to 120 seconds.

The table below lists the names of the 17 neurons, the time windows for which statistical significance as outlined in Section 2.3 was found, delay, duration, pattern change type (as seen on the visualisation of averages), and the type and functions of the neuron.

Name	Time window(s)	Delay (sec)	Duration (sec)	Pattern change type	Type and functions (some irrelevant functions not listed)
URADL	10	10	90	Rapid decrease followed by gradual increase back	Motor neuron for head muscles; may have sensory functions (47)
IL2R	20	20	20	Unclear	Sensory neuron; regulates nictation (40) (dispersal behaviour and survival strategy under harsh conditions (42))
ASHL	20	0	20	Unclear	Sensory neuron; **responds to noxious stimuli and leads to avoidance responses**, senses light and leads to reversals if a flash of light is focused on the head of a forwards-moving worm (34), **known connections to heat avoidance behaviour** (21)
FLPR	20	0	20	Rapid increase followed by gradual decrease back	Sensory neuron; **senses noxious temperatures and leads to a reflex-like escape response** (36)

NSML	10, 20	0, 0	120, 120	Rapid increase followed by gradual decrease back	Motor and sensory neuron; functions related to sensing food (41)
RMGL	5	10	70	Rapid increase followed by gradual decrease back	Interneuron; integrates signals from various sensory neurons; functions related to social behaviour (45)(46)
MCL	1	10	86	Rapid increase followed by gradual decrease back, followed by sustained decrease, followed by gradual increase back	Motor and possibly sensory neuron; controls the frequency of pharyngeal pumping (feeding behaviour indicating food intake) (42)
AINR	5	0	25	Rapid decrease followed by rapid increase followed by gradual decrease back	Interneuron; function unknown (49)
AWBL	20	0	40	Rapid increase followed by gradual decrease back	Sensory neuron; leads to odour avoidance (50); senses light and leads to a reversal when light is focused on the head of a forwards-moving worm (52); **known connections to heat avoidance behaviour** (21)
AVDL	5, 10, 20	0, 0, 0	105, 40, 60	Rapid increase followed by gradual decrease back	Interneuron; **drives backwards movement** (32)

AVER	5	0	75	Rapid decrease followed by rapid increase back	Interneuron; **drives backwards movement** (33)
ASGR	5	10	5	Unclear	Sensory neuron; controls (inhibits) entry into dauer stage (39), which is a larval stage which C.Elegans enter when conditions are too harsh for growth and reproduction (51)
ADAL	10, 20	0, 0	120, 120	Rapid increase followed by gradual/slow decrease back, followed by sustained decrease	Interneuron; function unknown (48)
ADLL	5	0	45	Unclear	Sensory neuron; “transmit information about **aversive stimuli in the environment** to a circuit that is responsible for aggregation, rapid locomotion, and food bordering behavior” (37)(38)
RIVL	1, 10	1, 0	96, 120	Rapid decrease followed by gradual increase, followed by sustained increase, followed by unclear decrease back	Interneuron and motor neuron; initiates local search behaviour for food which consists of reversals and turns (44)

AUAL	20	0	20	Rapid increase followed by gradual decrease back	Interneuron; regulates social and feeding behaviour (43)
I3	10	0	80	Rapid decrease followed by rapid increase	Interneuron and possibly sensory neuron; function unknown (35)

Source of neuron types: Individual neuron pages from (9)

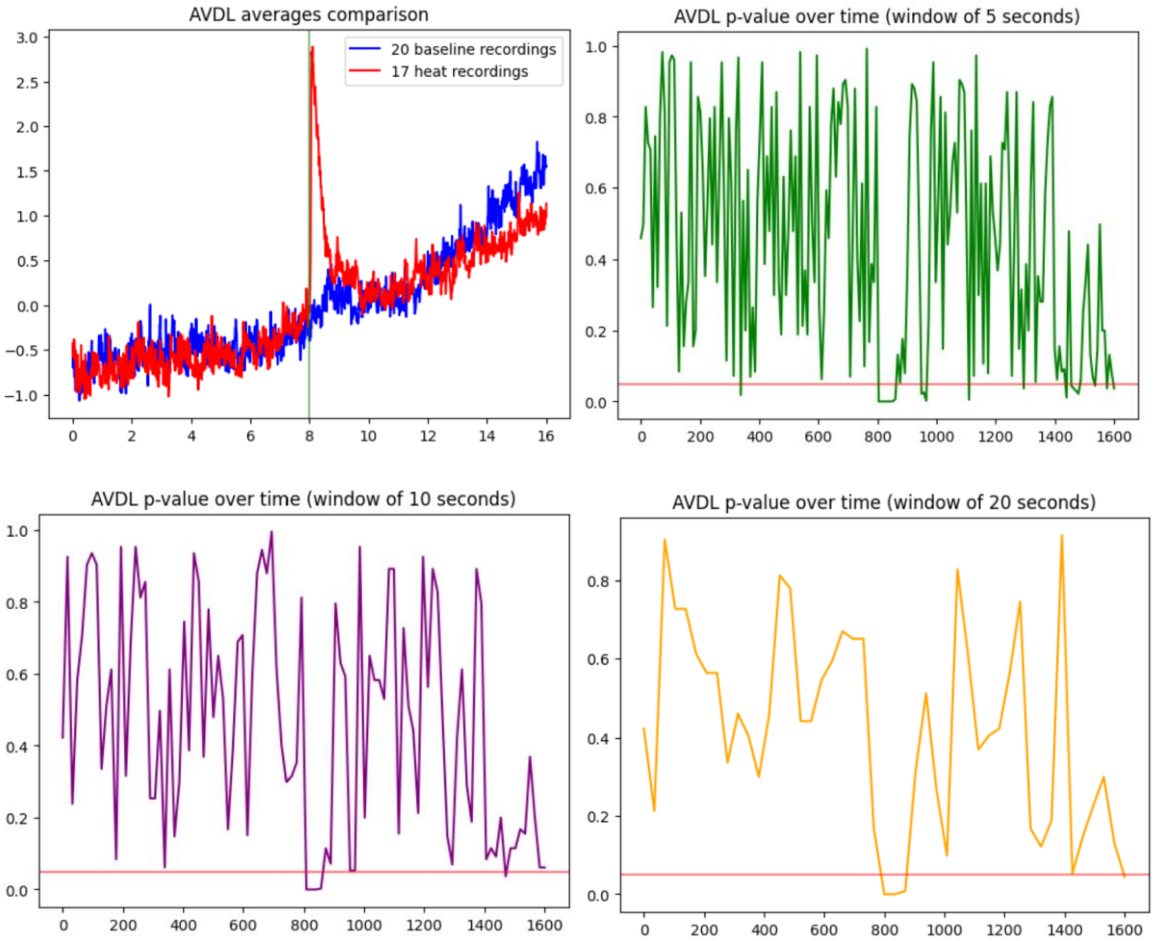


Fig.6: Top-left: Visualisation of averages of heat (red) vs baseline (blue) datasets for neuron "AVDL", where a clear change in pattern in the graph of averages is visible. Green vertical line shows when the heat stimuli was introduced at T=8 minutes. X-axis represents time (in minutes), and y-axis represents the strength of neural signal (in the number of standard deviations). Top-right and bottom: Visualisations of p-value over time for neuron "AVDL", for three window sizes which passed the statistical significance criteria. X-axis represents time (in time steps, where 100 time steps = 1 minute), and y-axis represents the p-value. The red horizontal line shows where p-value = 0.05. The p-value drops below 0.05 for all 3 windows

right after the introduction of the heat stimuli at T = 8 minutes (denoted by “800” on the x-axis).

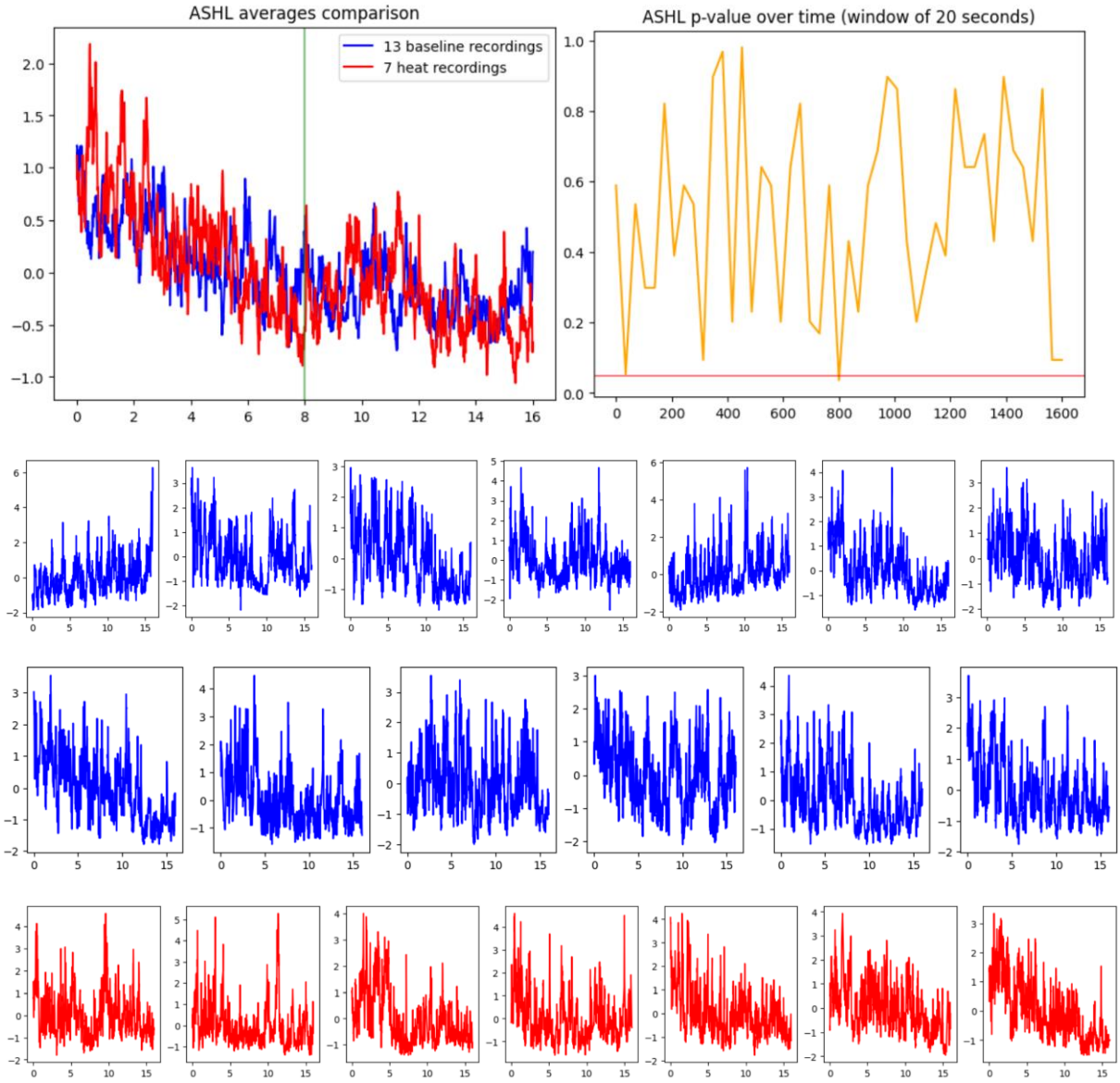


Fig.7: Top-left: Visualisation of averages of heat (red) vs baseline (blue) datasets for neuron “ASHL”, where there is no clear change in the pattern of averages after the stimulus is introduced (green vertical line). Top-right: Visualisations of p-value over time for neuron “ASHL”, for one window size which passed the statistical significance criteria. The p-value drops below 0.05 only at T=8 minutes, and gets close to this value once before the stimulus is introduced. Bottom 3 rows: individual measurements in the dataset for this neuron used to calculate the averages.

To see all visualisations of p-values and averages over time as well as the delay and duration of each identified time window, see Appendix D.

Chapter 4
Discussion

4.1 Visualisation, averages, and difference within group vs between groups

Taking an average across all individual measurements in heat versus baseline datasets appears to “smooth out” the differences between individual measurements and show a general trend they all follow. As an example in Fig.3, both heat and baseline datasets appear to follow the same pattern of gradually declining before minute 8, which is not clearly visible from graphs of individual measurements. At minute 8, there is a clear immediate rise in signal in the average of heat datasets. Some of the individual heat datasets appear to have a slight spike around minute 8, but this is not as clearly visible as it is with the average.

This “smoothing out” property might explain the discrepancy between the normalised difference within group and normalised difference between group averages, where the former tends to be about twice as high as the latter. Due to the discrepancy, we’re not able to simply compare the two normalised differences to e.g. identify neurons for which the difference within group is low and difference between groups is high (which would identify low-variability neurons that are likely to have a statistically significant difference between heat and baseline datasets). Future research might look further into methods of comparing the differences within group versus between groups.

For the neurons which appear to be high-variability neurons (due to having a high difference between groups in the timeframe 6 – 10 minutes but not showing a clear change in pattern around minute 8 on the graph of averages), one might expect the normalised difference within groups to be higher than low-variability neurons. However, for the neurons included in the visualisation that fit this criteria (“RIMR” and “RIML”, see Fig.4), this doesn’t appear to be the case. The averages of baseline and heat are visually clearly different from each other even before the stimulus (indicating a potentially high-variability neuron), however the difference within groups is similar or lower than the difference of neurons with visually very similar averages of baseline and heat before the stimulus e.g. “ADAL” neuron in Fig.3. This might be due to the “ADAL” neuron having more outliers and a wider range of values in the individual measurements (with some individual measurements reaching values as high as 15 with a likely outlier, and generally reaching values between -2 and 4 not considering outliers) while still following the same general trend which shows up in the average, whereas “RIMR” and “RIML” neurons appear to have less outliers and a smaller range between -1 and 3

while possibly not following a general pattern, and rather just randomly oscillating within this range. Therefore, it's possible that the average "smooths out" these outliers and large ranges, but they play a role in inflating the average within-group difference. It's also possible that at least some of the neurons which have a high difference between groups in the timeframe 6 – 10 minutes but don't show a clear change in pattern around minute 8 on the graph of averages are not high-variability neurons as we previously assumed, but rather they are neurons that don't follow a general trend over time. Furthermore, it's possible that the significant visual differences between the group averages in "RIML" and "RIMR" neurons are due to them having a smaller number of individual measurements than the "ADAL" neuron which leads to less uniform averages, rather than a high variability of the neuron itself.

As mentioned in Section 3.2, 64 out of 109 considered neurons show the highest difference between the averages of heat and baseline datasets to be in the timeframe 6 – 10 minutes, which includes the time during and shortly after the stimulus. This can be interpreted as the noxious heat stimulus having a wide effect on the nervous system of the worm, likely due to the stimulus being indicative of danger.

4.2 Finding statistically significant neurons and p-value as a function of time

Despite the strict statistical significance criteria, a relatively large number of 17 significant neurons is identified. This might further indicate that noxious heat stimuli elicit a wide response from the nervous system of the worm.

6 out of 17 neurons have known functions to either heat avoidance behaviour, other avoidance behaviour, or backwards movement. 5 of these neurons ("FLP", "AVER", "AVDL", "ASHL", "AWBL") match the neurons previously identified as part of a heat avoidance circuit as explained in Section 1.1.2. This means that our methods are successful in identifying at least some of the neurons directly involved in the heat avoidance circuit. The remaining neuron "ADLL" has been linked to other avoidance behaviour (38). It's possible the results of our work indicate that another function of "ADLL" is detecting aversive heat stimuli from the environment, especially because "ADLL" is a sensory neuron with direct connections to the outside environment (29), and being on the "front lines" could mean that its statistical significance is more likely due to a direct response to the stimulus rather than the neuron being indirectly affected by upstream activity of other neurons.

Of the remaining 9 neurons, some have a function that might be related to heat avoidance. “IL2R” neuron regulates nictation which is a survival strategy under harsh conditions (40)(42), and “ASGR” controls entry into a larval stage which C. Elegans enter when conditions are too harsh for reproduction or growth (39)(51). As noxious heat stimuli can be indicative of harsh conditions, it’s possible that the response of those neurons is directly linked to it. Similarly, the response of some of the remaining neurons responsible for social behaviour might be due to the stress of the noxious stimulus inhibiting social functions. This is speculative, and future work might involve lab-based approaches to further investigate the identified neurons. This work has identified 17 neurons that might be directly linked to heat avoidance behaviour and that can serve as a “shortlist” for future lab-based experiments. Future lab-based research could involve inhibiting the function of individual or a combination of those neurons to investigate how the heat response or other behaviour changes.

As mentioned in Section 3.2, 64 out of 109 considered neurons show the highest difference between group averages to be in the timeframe 6 – 10 minutes. However, a much smaller number of 17 neurons pass the statistical significance criteria. This is likely due to many neurons being highly variable in general and showing p-values < 0.05 for many time windows other than the time windows immediately following the stimulus. It’s also possible that the statistical significance criteria that we use is too strict and some of the neurons which are in fact involved in the heat avoidance response were eliminated – it’s possible that some of those neurons are high-variability neurons, or that they have multiple functions beyond heat avoidance which led them to be significantly active at other times, which then led to p-values < 0.05 for time windows before the stimulus. This might explain why not all of the known neurons responsible for heat avoidance behaviour have been identified. Future research might consider methods of identifying high-variability neurons, detecting statistically significant activity in high-variability neurons, and differentiating between high-variability neurons and neurons with multiple functions.

Most of the 17 neurons only show statistical significance in only one of the potential time windows. Some of the neurons show it in two windows, with “AVDL” being the only neuron with three windows. As “AVDL” has a very clear visual change in pattern and it’s one of the neurons previously identified to be part of the heat avoidance circuit, it’s possible that the higher number of windows corresponds to a more clear change in pattern. Furthermore, as most of the neurons only show significance in one window, it’s possible that some significant neurons were missed in this method due to not testing a window size that would show a significance for them. Alternatively, if a neuron would show significance in only one of many potential time windows, it’s possible that the neuron is not significant and it just happened to return a significant result for this one window due to chance. Future work could involve

investigating more potential time windows as well as investigating whether certain sizes of time windows are more indicative of significance than others.

Most of the 17 neurons show a clear visual change in pattern, for example the neuron “AVDL” as visible in Figure 6. Some neurons don’t appear to have a clear change in pattern while analysing the graph of averages, such as neuron “ASHL” (Fig. 7). However, this neuron passes the strict significance criteria, and from the graph of p-value over time it’s visible that the p-value clearly drops below 0.05 only at the time of the stimulus. The p-value only gets close to 0.05 at one time before the stimulus. It’s possible that the KS test returns a significant result due to a change in pattern not visible on the graph of averages, such as was illustrated in Figure 1. Future work could involve investigating the individual measurements of the 17 identified neurons in order to better determine the type of the pattern change.

A limitation of this work is that we can’t distinguish whether the identified neurons are directly involved in the heat avoidance circuit, or whether they are simply downstream neurons affected by the signals sent from other neurons upstream. We have shown that the neurons show a statistically significant change in their activity immediately after the noxious heat stimulus while not showing such a significant change in the time before, however this doesn’t constitute a proof of the neurons being directly involved in creating the heat avoidance behaviour. As noxious heat stimulus leads to a wide response in the worms’ nervous system, many neurons will be inadvertently affected through their connections to other neurons that are more directly involved in the behaviour. Future work could attempt to distinguish between direct versus downstream neurons, for example by investigating the delay between the stimulus and the change in neural activity.

The delay for all neurons with known connections to heat avoidance behaviour is 0, and delays for other neurons range from 0 to 20 seconds. No delay might indicate that the neuron is directly responding to the stimulus, and longer delays might indicate that it’s a downstream response or even that it’s a response not related to the stimulus. Future work can involve investigating what an expected delay for neurons is, dependent on what is known about the chemical and anatomical properties that affect the speed of neural signal propagation.

The duration ranges from a single time window to 120 seconds. Sensory neurons tend to have a shorter duration, which could be due to them only transmitting the immediate sensory information to other neurons connected to them. Motor neurons tend to have a longer duration, which might indicate that the muscles and movement of the animal is affected for a

significant length of time after the stimulus, for example by the worm moving backwards or simply being in a state of “stress”.

As this work is not lab-based but rather only based on investigating a dataset with statistical methods, it can be applied to neural recordings of animals other than C. Elegans such as fruit flies, mice, chimpanzees, and humans. The method could be applied to recordings in any format which represents the strength of neural signals over time, therefore it could be used with e.g. EEG recordings. Future work could involve testing whether the method returns satisfying results for more complex animals, and applying the method to identify statistically significant neurons linked to behaviours that they exhibit.

List of References

(1) University of Minnesota. *What is Caenorhabditis elegans and why work on it?*. [Online]. [no date]. [Accessed April 2024]. Available from: https://cgc.umn.edu/what-is-c-elegans

(2) National Library of Medicine. Search for term “C.Elegans”. [Online]. [Accessed April 2024]. Available from: https://pubmed.ncbi.nlm.nih.gov/?term=C.Elegans

(3) Hart, A.C. Behaviour. In: Ambros, A. ed. *WormBook.* [Online]. The C. elegans Research Community, 2006. [Accessed April 2024]. Available from: http://www.wormbook.org/chapters/www_behavior/behavior.html

(4) Riddle, D.L., Blumenthal, T., Meyer, B.J., et al. Section I - The Biological Model. In: *C. elegans II*. [Online]. 2nd edition. Cold Spring Harbor Laboratory Press, 1997. [Accessed April 2024]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK20086/

(5) Richmond, J. Synaptic function. In: Jorgersen, M.E. and Kaplan, J.M. ed. *WormBook*. [Online]. The C. elegans Research Community, 2007. [Accessed April 2024]. Available from: http://www.wormbook.org/chapters/www_synapticfunction/synapticfunction.hml

(6) Schafer, W.R. Neurophysiological methods in C. elegans: an introduction. In: Ambros, V. ed. *WormBook.* [Online]. The C. elegans Research Community, 2006. [Accessed April 2024]. Available from: http://www.wormbook.org/chapters/www_intromethodsneurophys/intromethodsneurophys.html

(7) Naudin, L., Laredo, J.L.J., Liu, Q., and Corson, N. Systematic generation of biophysically detailed models with generalization capability for non-spiking neurons. *PLOS ONE.* 2022.

(8) Jiang, J., Su, Y., Zhang, R., Li, H., Tao, L., and Liu, Q. C. elegans enteric motor neurons fire synchronized action potentials underlying the defecation motor program. *Nature Communications.* 2022, **13**, Article number: 2783.

(9) Altun, Z. *Individual Neurons*. [Online]. [no date]. [Accessed April 2024]. Available from: https://www.wormatlas.org/neurons/Individual%20Neurons/Neuronframeset.html

(10) Hobert, O. Development of left/right asymmetry in the Caenorhabditis elegans nervous system: From zygote to postmitotic neuron. *Genesis.* 2014, **52**(6), pp.528-543.

(11) Cook, S.J., Jarrell, T.A., Brittin, C.A. et al. Whole-animal connectomes of both Caenorhabditis elegans sexes. *Nature*. 2019, **571**(7763), pp.63-71.

(12) Portegys, T.E. Training sensory–motor behavior in the connectome of an artificial C. elegans. *Neurocomputing*. 2015, **168**, pp.128-134.

(13) Randi, F., Sharma, A.K., Dvali, S., and Leifer, A.M. Neural signal propagation atlas of Caenorhabditis elegans. *Nature*. 2023, **623**(7986), pp.406-414.

(14) Atanas, A.A., Kim, J., Wang, Z., et al. Brain-wide representations of behavior spanning multiple timescales and states in C. elegans. *Cell*. 2023, **186**(19), pp.4134-4151.

(15) Kyriakou, E., Taouktsi, E., Syntichaki, P. *The Thermosensory Circuit of the Nematode Caenorhabditis elegans: History*. [Online]. [no date]. [Accessed April 2024]. Available from: https://encyclopedia.pub/entry/history/show/85889

(16) Ikeda, M., Nakano, S., Giles, A.C. et al. Context-dependent operation of neural circuits underlies a navigation behavior in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America.* 2020, **117**(11), pp.6178-6188.

(17) Kuhara, A., Okumura, M., Kimata, T. et al. Temperature sensing by an olfactory neuron in a circuit controlling behavior of C. elegans. *Science*. 2008, **320**(5877), pp. 803-807.

(18) Kyriakou, E., Taouktsi, E., Syntichaki, P. The Thermal Stress Coping Network of the Nematode Caenorhabditis elegans. *International Journal of Molecular Sciences*. 2022, **23**(23).

(19) Liu, S., Schulze, E., and Baumeister, R. Temperature- and Touch-Sensitive Neurons Couple CNG and TRPV Channel Activities to Control Heat Avoidance in Caenorhabditis elegans. *PLOS ONE*. 2012, **7**(3).

(20) Marques, F., Saro, G., Lia, A-S. et al. Identification of avoidance genes through neural pathway-specific forward optogenetics. *PLOS GENETICS*. 2019, **15**(12).

(21) Ward, A., Liu, J., Feng, Z., Xu, X.Z.S. Light-sensitive neurons and channels mediate phototaxis in C. elegans. *Nature Neuroscience*. 2008, **11**(8), pp. 916-922.

(22) Chiu, H., Robie, A.A., Branson, K.M. et al. [Pre-print]. Cell type-specific contributions to a persistent aggressive internal state in female Drosophila. *eLife*. 2023.

(23) Kim, H-Y. Analysis of variance (ANOVA) comparing means of more than two groups. *Restorative Dentistry and Endodontics.* 2014, **39**(1), pp.74-77.

(24) Darling, D.A. The Kolmogorov-Smirnov, Cramer-von Mises Tests. *The Annals of Mathematical Statistics*. 1957, **28**(4), pp. 823-838.

(25) National Institute of Standards and Technology. *Kolmogorov Smirnov Two Sample*. [Online]. 2001. [Accessed April 2024]. Available from: https://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/ks2samp.htm

(26) GraphPad. *Interpreting results: Kolmogorov-Smirnov test*. [Online]. [no date]. [Accessed April 2024]. Available from:

https://www.graphpad.com/guides/prism/latest/statistics/interpreting_results_kolmogorov-smirnov_test.htm

(27) Rice University. *Critical Values for the Two-sample Kolmogorov-Smirnov test (2-sided).* [Online]. [no date]. [Accessed April 2024]. Available from: https://sparky.rice.edu/astr360/kstest.pdf

(28) Andrade, C. The P Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives. *Indian Journal of Psychological Medicine*. 2019, **41**(3), pp.210-215.

(29) White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. The Structure of the Nervous System of the Nematode Caenorhabditis elegans. In: Boulin, T. Web adaptation. The Mind of a Worm. (Online). Web adaptation: *WormAtlas.* 2001. Original: *Philosophical Transactions of the Royal Society of London.* 1986, **314**(1165). (Web adaptation accessed April 2024). Available from: https://www.wormatlas.org/MoW_built0.92/cells/adl.html

(30) Lanzante, J.R. Testing for differences between two distributions in the presence of serial correlation using the Kolmogorov–Smirnov and Kuiper's tests. *International Journal of Climatology.* 2021, **41**(14), pp.6314-6323.

(31) PennState Eberly College of Science. *Autocorrelation and Time Series Methods.* [Online]. [no year]. [Accessed April 2024]. Available from: https://online.stat.psu.edu/stat462/node/188/

(32) Piggott, B.J., Liu, J. Feng, Z., Wescott, S.A., and Xu, X.Z.S. The Neural Circuits and Synaptic Mechanisms Underlying Motor Initiation in C. elegans. *Cell*. 2011, **147**(4), pp.922-933.

(33) WormAtlas. *AVEL, AVER*. [Online]. [no year]. [Accessed April 2024]. Available from: https://www.wormatlas.org/neurons/Individual%20Neurons/AVEframeset.html

(34) WormAtlas. *ASHL, ASHR.* [Online]. [no year]. [Accessed April 2024]. Available from: https://www.wormatlas.org/neurons/Individual%20Neurons/ASHframeset.html

(35) WormAtlas. *I3*. [Online]. [no year]. [Accessed April 2024]. Available from: https://www.wormatlas.org/neurons/Individual%20Neurons/I3frameset.html

(36) Liu, S., Schulze, E., and Baumeister, R. Temperature- and Touch-Sensitive Neurons Couple CNG and TRPV Channel Activities to Control Heat Avoidance in Caenorhabditis elegans. *PLOS ONE*. 2012, **7**(3).

(37) de Bono, M., Tobin, D.M., Davis, M.W., Avery, L., and Bargmann, C.I. Social feeding in Caenorhabditis elegans is induced by neurons that detect aversive stimuli. *Nature*. 2002, **419**(6910), pp.899-903.

(38) WormAtlas. *ADLL, ADLR.* [Online]. [no year]. [Accessed April 2024]. Available from: https://www.wormatlas.org/neurons/Individual%20Neurons/ADLframeset.html

(39) Bargmann, C.I., Horvitz, H.R. Control of Larval Development by Chemosensory Neurons in Caenorhabditis elegans. *Science.* 1991, **251**(4998), pp.1243-1246.

(40) Lee, H., Choi, M, Le, D., et al. Nictation, a dispersal behavior of the nematode Caenorhabditis elegans, is regulated by IL2 neurons. *Nature Neuroscience*. 2011, **15**(1), pp. 107-112.

(41) Albertson, D.G., Thompson, J.N. The pharynx of Caenorhabditis elegans. *Philosophical Transactions of the Royal Society B Biological Sciences.* 1976, **275**(938), pp.299-325.

(42) Riddle, D.L., Blumenthal, T., Meyer, B.J., et al. Section II - Feeding. In: *C. elegans II*. [Online]. 2nd edition. Cold Spring Harbor Laboratory Press, 1997. [Accessed April 2024]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK20086/

(43) Coates, J.C. and de Bono, M. Antagonistic pathways in neurons exposed to body fluid regulate social feeding in Caenorhabditis elegans. *Nature*. 2002, **419**(6910), pp.925-929.

(44) Gray, J.M., Hill, J.J., and Bargmann, C.I. A circuit for navigation in Caenorhabditis elegans. *The Proceedings of the National Academy of Sciences.* 2005, **102**(9), pp.3184-3191.

(45) Macosko, E.Z., Pokala, N., Feinberg, E.H. et al. A Hub-and-Spoke Circuit Drives Pheromone Attraction and Social Behavior in C. elegans. *Nature*. 2009, **458**(7242), pp.1171-1175.

(46) Lockery, S.R. A social hub for worms. *Nature*, 2009, **458**(7242), pp.1124-1125.

(47) WormAtlas. *URADL, URARD, URAVL, URAVR*. [Online]. [no year]. [Accessed April 2024]. Available from: https://www.wormatlas.org/neurons/Individual%20Neurons/URAframeset.html

(48) WormAtlas. *ADAL, ADAR.* [Online]. [no year]. [Accessed April 2024]. Available from: https://www.wormatlas.org/neurons/Individual%20Neurons/ADAframeset.html

(49) WormAtlas. *AINL, AINR.* [Online]. [no year]. [Accessed April 2024]. Available from: https://www.wormatlas.org/neurons/Individual%20Neurons/AINframeset.html

(50) Troemel, E.R., Kimmel, B.E., Bargmann, C.I. Reprogramming Chemotaxis Responses: Sensory Neurons Define Olfactory Preferences in C. elegans. *Cell*. 1997, **91**(2), pp.161-169.

(51) Xantha, K. Working with dauer larvae. In: Hobert, O. ed. *WormBook*. [Online]. The C. elegans Research Community, 2018. [Accessed April 2024]. Available from: http://www.wormbook.org/chapters/www_dauerlarvae/dauerlarvae.html#ftn.d0e29

(52) WormAtlas. *AWBL, AWBR*. [Online]. [no year]. [Accessed April 2024]. Available from: https://www.wormatlas.org/neurons/Individual%20Neurons/AWBframeset.html

(53) Wittenburg, N, Baumeister, R. Thermal avoidance in Caenorhabditis elegans: An approach to the study of nociception. *Proceedings of the National Academy of Sciences of the United States of America.* 1999, **96**(18), pp.10477-10482.

(54) Gordus, A., Pokala, N., Levy, S., Flavell, S.W., and Bargmann, C.I. Feedback from Network States Generates Variability in a Probabilistic Olfactory Circuit. *Cell*. 2015, **161**(2), pp.215-227.

Appendix A
Self-appraisal

A.1 Critical self-evaluation, personal reflection and lessons learned

This was the most challenging university project I have completed. I chose a difficult topic that I knew would push me to learn new practical skills that I haven't used before. I initially assumed that the project would involve investigating various pattern-detecting algorithms, but it turned out to be very focused on statistical methods instead. This was a challenge, because I haven't had any experience with statistics before this project, and I had to individually learn the relevant topics from scratch. However, I believe that I have been successful in this, and I have been able to understand the Kolmogorov-Smirnov test to the extend of being able to apply it as part of a method that I haven't seen used in this specific way anywhere else. What helped the most was seeing a relatively similar method in a published paper which gave me the confidence that this method can be applied successfully to this type of data, and that it doesn't break some of the underlying assumptions of the KS test. If I had not seen the method in the 2023 research of Randi, F., Sharma, A.K., Dvali, S. and Leifer, A.M., I would have likely used a simpler or more limited method such as ANOVA. In this project I have learned the importance of being familiar with the literature and reading around a subject, because a method mentioned in passing in one paper could potentially be a foundational method used in my work, as it happened with the KS test.

Admittedly, I have not been able to fully answer the posed question of "which neurons are responsible for heat avoidance behaviour in C. Elegans". I set out hoping that I could identify the heat avoidance circuit, but through the course of the work I've come to realise that even if I manage to identify statistically significant neurons, with the use of my current method it's not possible to tell whether those neurons are actually part of the circuit, or whether they're indirectly affected by the response of upstream neurons. More work would be needed in order to get closer to the answer to my question, for example considering connections between neurons and delays between responses of connected neurons. Unfortunately, due to time limitations and due to my limited knowledge, I was not able to include these as part of this project.

One of the first steps that I took as part of this project was to explore the datasets and visualise them, so that I can gain a contextual understanding of what I'm working with and try to look for any patterns that would give me an initial idea of what to focus on. Thanks to this, I was able to see that some of the averages of heat datasets have a clear change of pattern around the 8 minute mark. However, my supervisor pointed out examples where certain

changes of pattern would not be visible in the averages, for example as in Figure 1. Furthermore, I've realised that a visual change of pattern of the average doesn't constitute a proof of the neuron's significance – it might be visible to a human eye and it might be an indicator of something of interest, but another method that returns a quantifiable results is necessary in order to "prove" anything. This pushed me to explore more advanced methods such as ANOVA and the KS test, which return a p-value which is a generally accepted form of proof of statistical significance. From this, I have learned that even though something might visually seem obvious to a human eye, a quantifiable and rigorous method is necessary in order for the result to be accepted in the scientific community. I have also learned that investigating the data and visualising it in different ways can help to put the quantifiable results into context, though by itself it doesn't constitute a proof of statistical significance.

What surprised me about this project is that the background reading as well as planning and considering different aspects of my planned work took up significantly more time than coding the actual implementation in Python. Two parts of the background research took up the most time: identifying the appropriate method and ensuring that the method produces a result that makes sense in the context of the method's limitations and the dataset I'm using. I've considered many various potential approaches for identifying significant neurons, most of which would not fit quite right with the data or were not previously used in research for a similar purpose. As I'm only an undergraduate student without a background in research or statistics, I wanted to choose a method that has been used before for a similar purpose, because that would give me the confidence that it's a reasonable approach. Therefore, most of my time has been spent exploring various papers in search of an appropriate method. Once I identified a potential method, I had to learn about it from scratch and understand its scope and limitations, for example the underlying assumptions of the KS test which made it necessary to supply the data as an average over a time window, rather than as individual datapoints. I'm glad that I took the time to understand the method before jumping to implementing it, because it made the coding process relatively quick and easy.

Another reason for background research taking up the largest portion of time is due to the plethora of information and literature available about C. Elegans and statistical method. I tried to learn as much as possible to gain more context and understanding of my research problem, however I had to find the right point at which to stop in order to ensure that my project is completed on time. One of the reasons why I chose this challenging project is so that I can learn new topics and tools, and I'm very glad for having the opportunity to do so. This project pushed me to develop very valuable skills of statistical analysis, working with

large datasets, and reviewing literature, which I am certain will prove useful in my future career.

A.2 Legal, social, ethical and professional issues

A.2.1 Legal issues

This issue is not relevant to my project. I did not need permission to use any of the tools or data, as they are all publicly available. An example where legal issues could arise would be if recordings of human brains were used, which might introduce legal issues of privacy protections. However, to the best of my knowledge, there is no legislation restricting research on neural recordings of brains of worms.

A.2.2 Social and ethical issues

These issues are not relevant to my project. If the project involved recordings of human brain activity, there could be ethical issues associated with consent or social issues linked to ensuring that underrepresented groups are considered. However, the project only considers recordings of brains of worms and is not directly linked to research on humans, and therefore those issues are not present.

A.2.4 Professional issues

There were no professional issues with my project. I worked on the project individually, and there was no one else working on a similar project at the same time. The datasets used were publicly available and I didn’t need to ask for permission to use them.

Appendix B
External Materials

The external materials used were as follows:

- The datasets of worms' neural activity, downloaded from www.wormwideweb.org
- The KS test function ("ks_2samp") from Python SciPy library (as described here: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ks_2samp.html)
- Matplotlib Python module, used to generate the graphs of neural activity and p-value as a function of time

Appendix C

Graphs of averages of top 10 neurons

This appendix contains the visualisations of averages of top 10 neurons (ordered by the largest difference between the heat and baseline datasets during the timeframe of 6 – 10 minutes), the individual measurements used to calculate the averages, and the normalised values of the differences between and within groups.

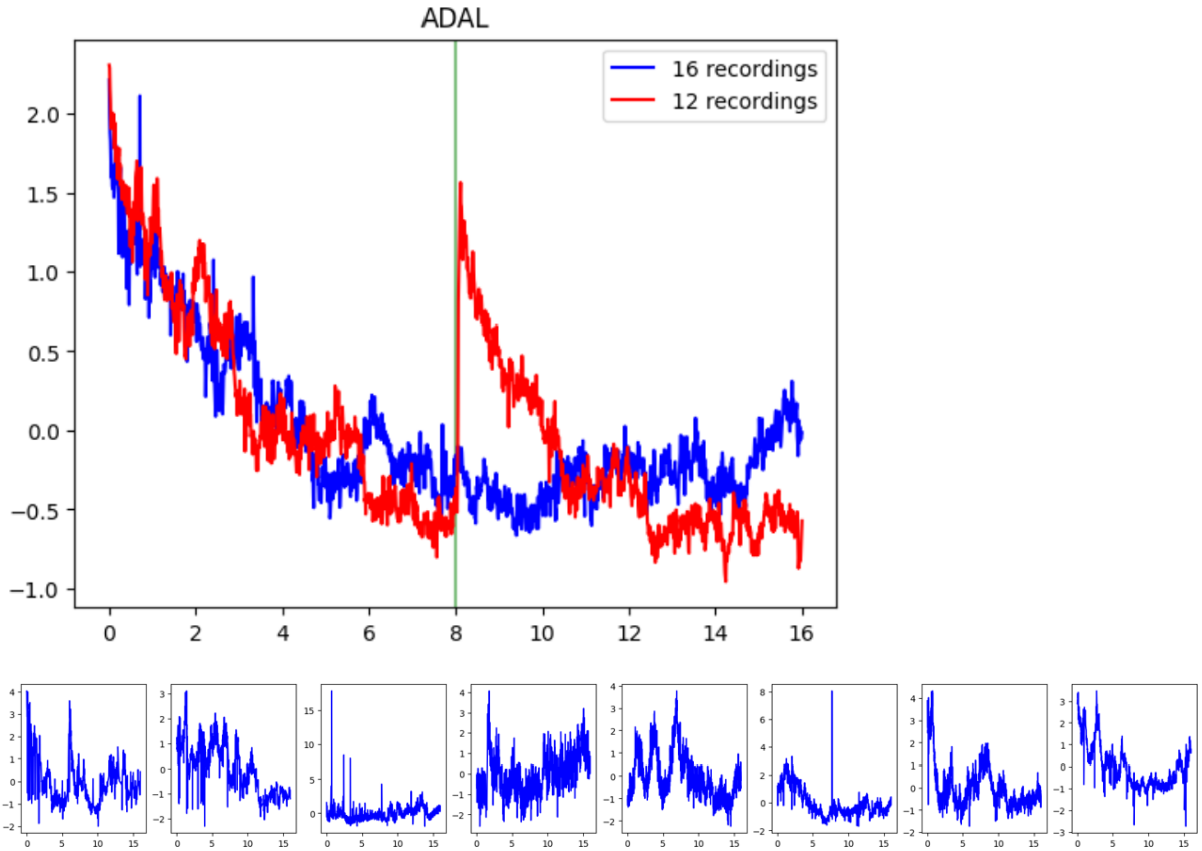
ADAL:

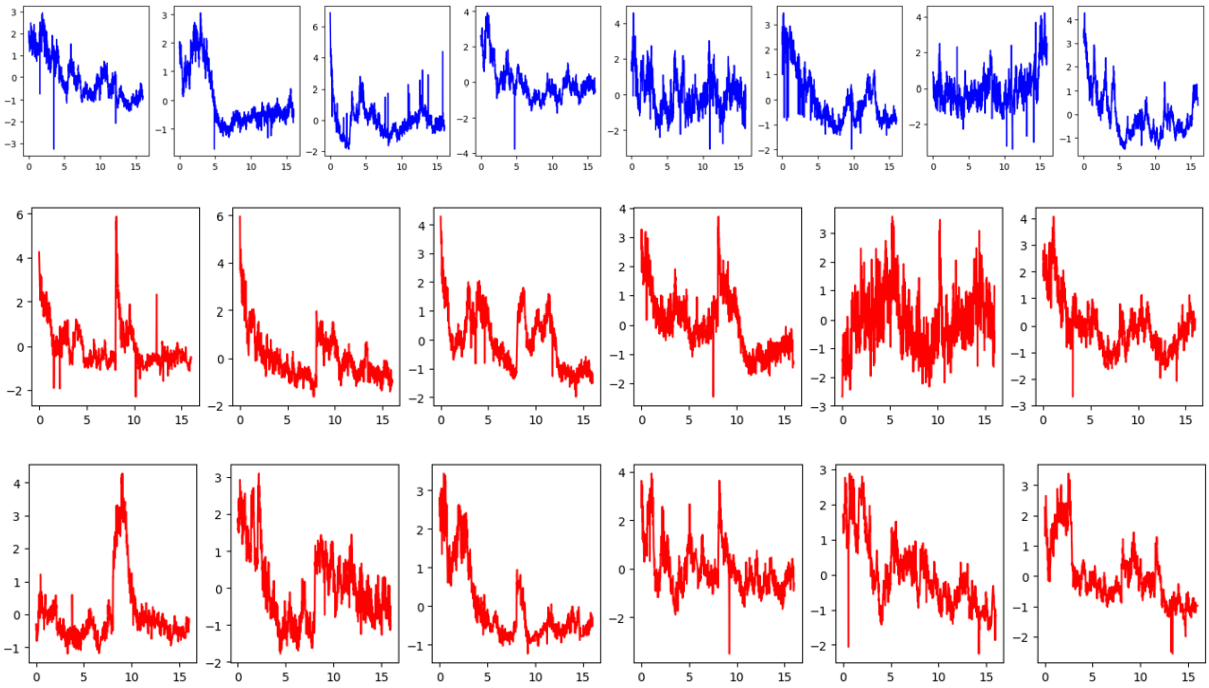
Absolute difference between avg heat and avg baseline (all time ranges normalised by dividing by the number of measurements in that time range):

- Total: 0.36938092860999083
- 6 mins - 10 mins: 0.6357640314354449
- After 8 mins: 0.47511146314001523

Average absolute difference within group (normalised):

- Baseline: 0.9461370268691266
- Heat: 0.8279549472239018





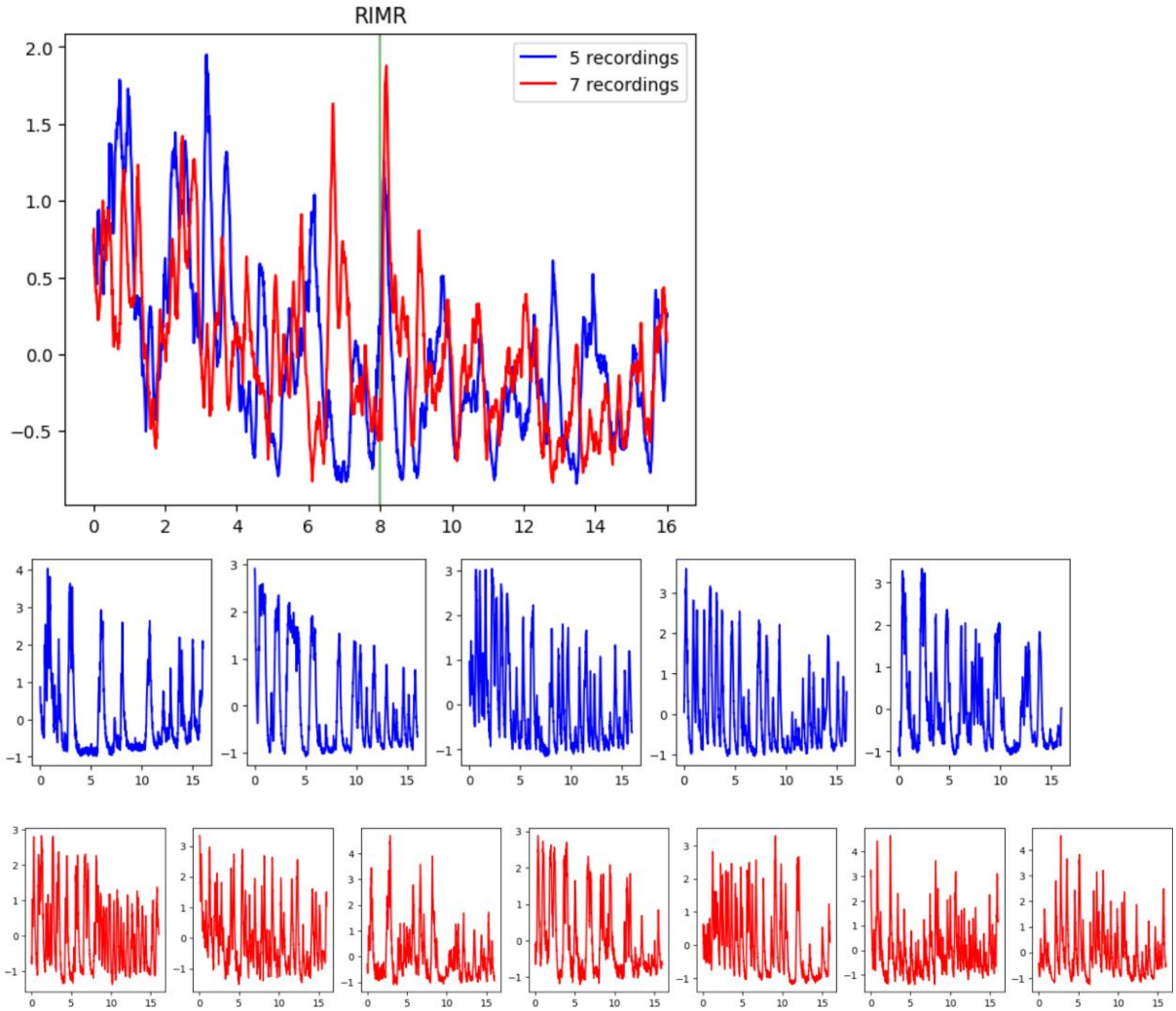
RIMR:

Absolute difference between avg heat and avg baseline (all time ranges normalised by dividing by the number of measurements in that time range):

- Total: 0.4662489870487432
- 6 mins - 10 mins: 0.6085925388689667
- After 8 mins: 0.34731064231262165

Average absolute difference within group (normalised):

- Baseline: 0.9419081186355691
- Heat: 1.0185651747375957



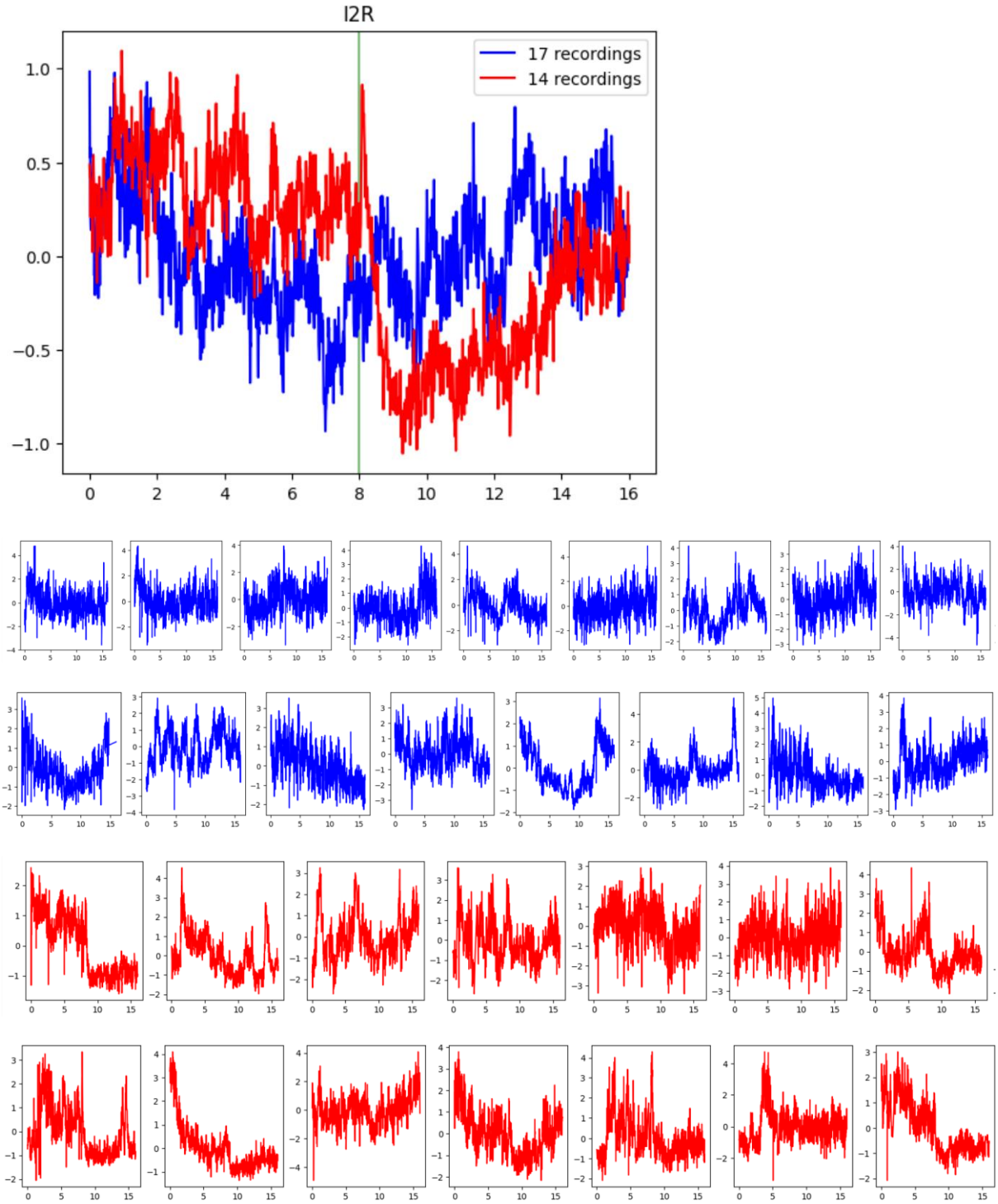
I2R:

Absolute difference between avg heat and avg baseline (all time ranges normalised by dividing by the number of measurements in that time range):

- Total: 0.47300047067722645
- 6 mins - 10 mins: 0.5455618213902548
- After 8 mins: 0.48870446635933673

Average absolute difference within group (normalised):

- Baseline: 1.092962287824133
- Heat: 0.9849223275062389



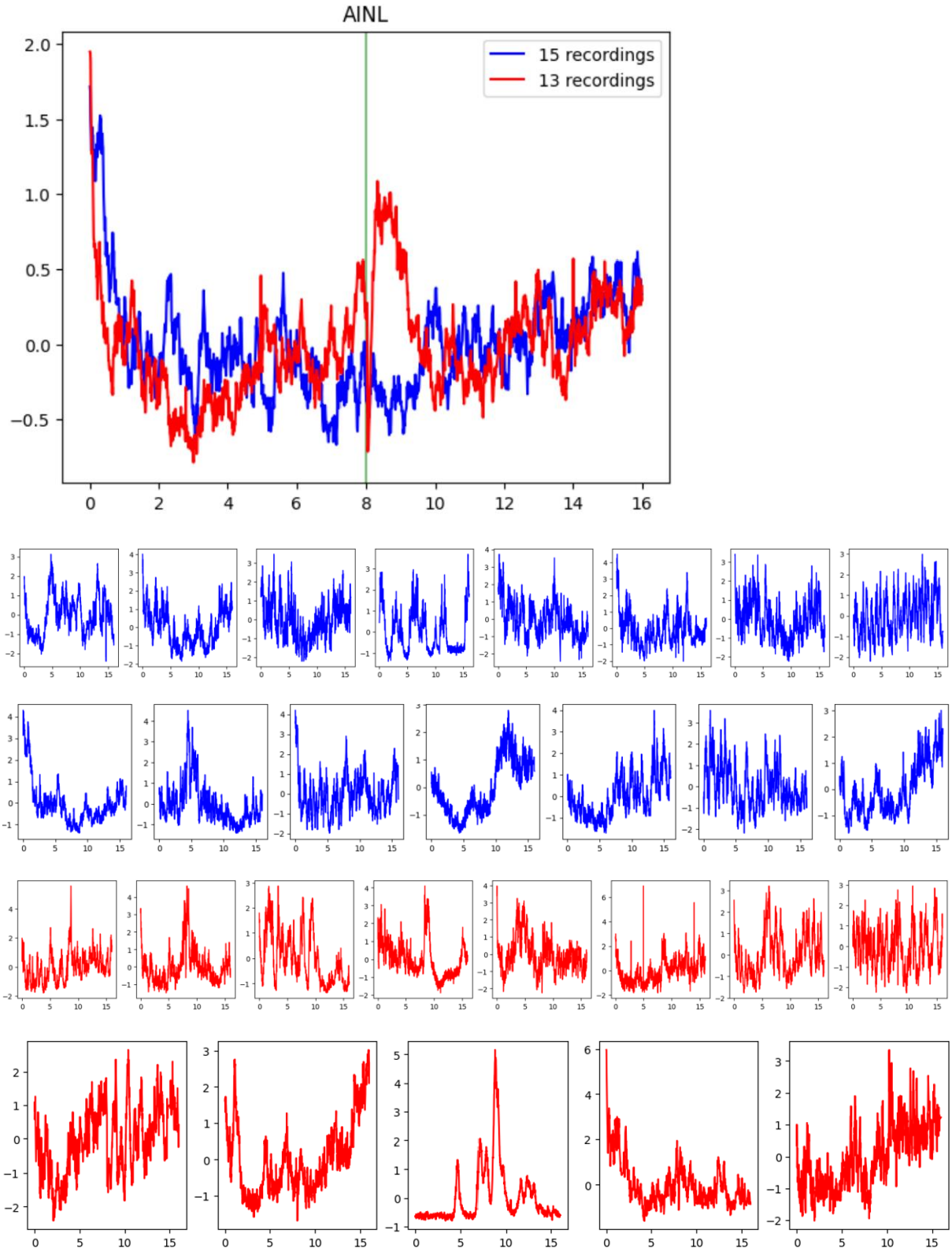
AINL:

Absolute difference between avg heat and avg baseline (all time ranges normalised by dividing by the number of measurements in that time range):

- Total: 0.3371962954134448
- 6 mins - 10 mins: 0.5374170381418755
- After 8 mins: 0.32900244329495704

Average absolute difference within group (normalised):

- Baseline: 1.0812873596459638
- Heat: 1.0673636767819445



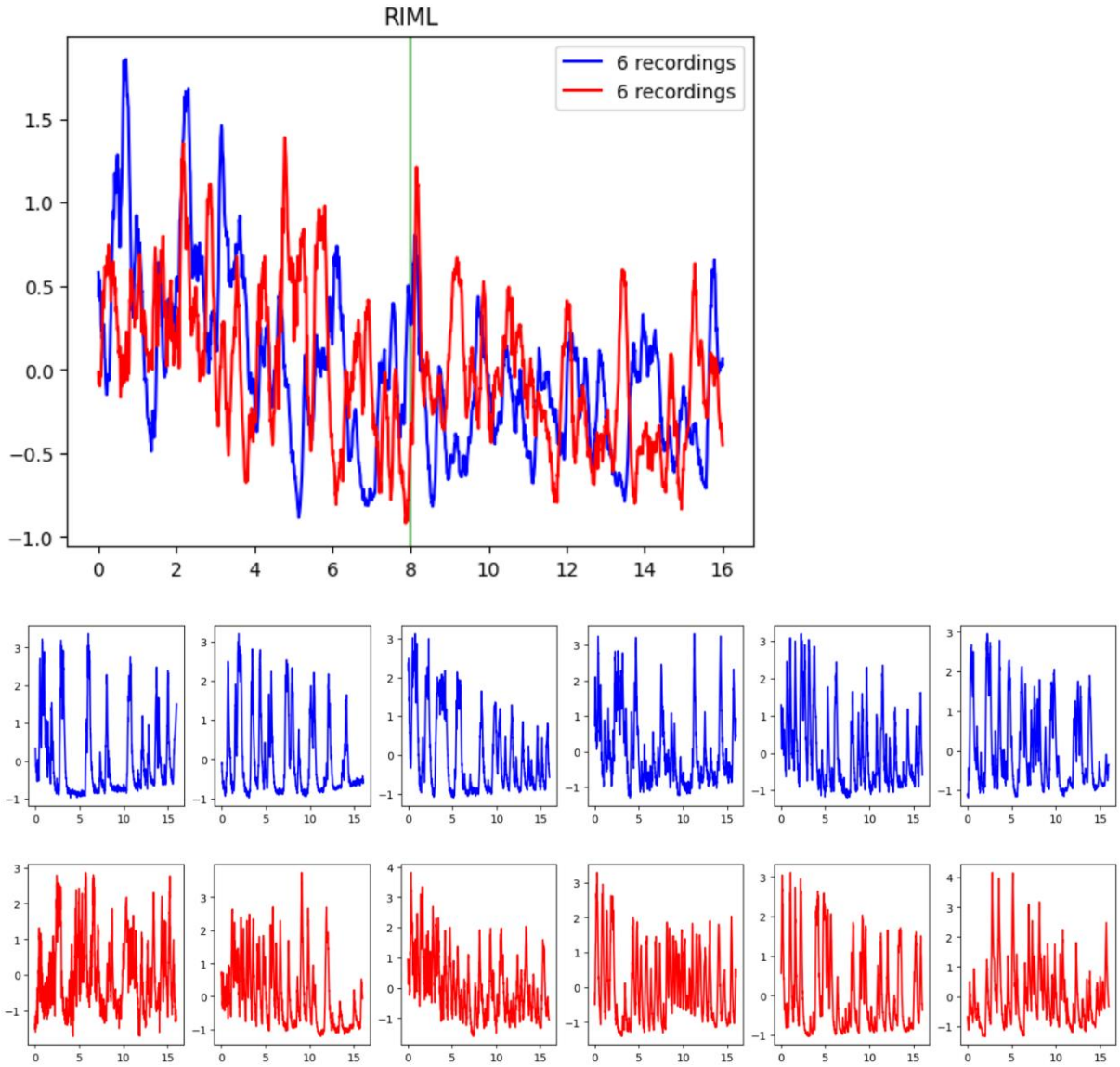
RIML:

Absolute difference between avg heat and avg baseline (all time ranges normalised by dividing by the number of measurements in that time range):

- Total: 0.48517122380549393
- 6 mins - 10 mins: 0.5353050680454731
- After 8 mins: 0.39935072019842993

Average absolute difference within group (normalised):

- Baseline: 0.9874212264403215
- Heat: 1.0800852913749996



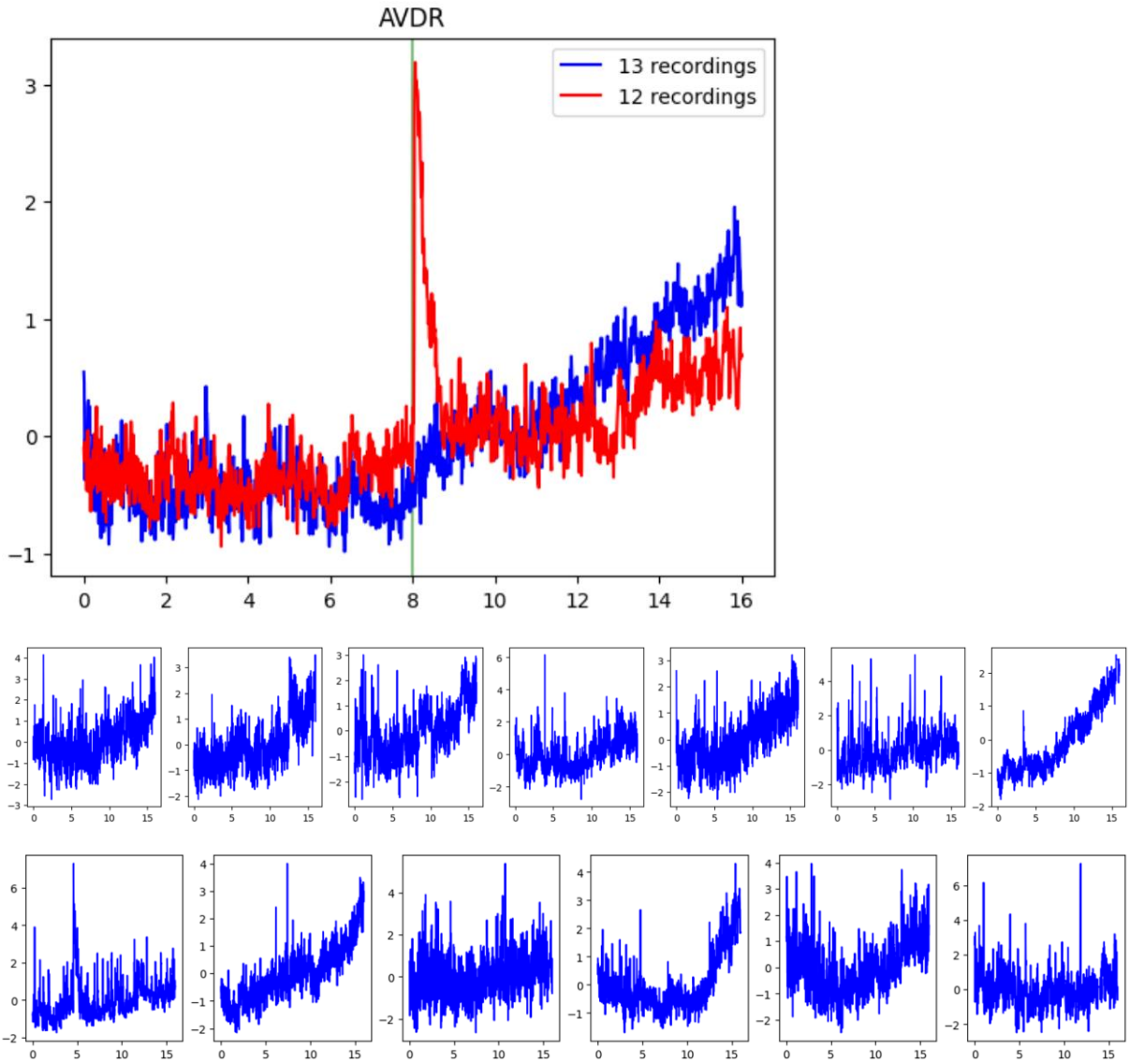
AVDR:

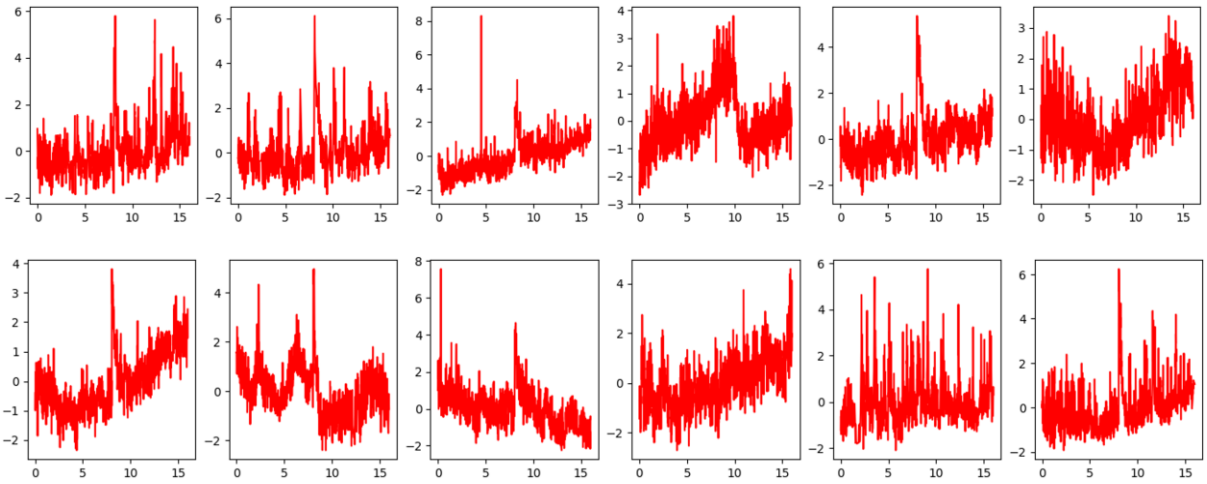
Absolute difference between avg heat and avg baseline (all time ranges normalised by dividing by the number of measurements in that time range):

- Total: 0.3824511985859299
- 6 mins - 10 mins: 0.5137185866540885
- After 8 mins: 0.5030986286610394

Average absolute difference within group (normalised):

- Baseline: 0.8628066926883347
- Heat: 0.9717031374406561





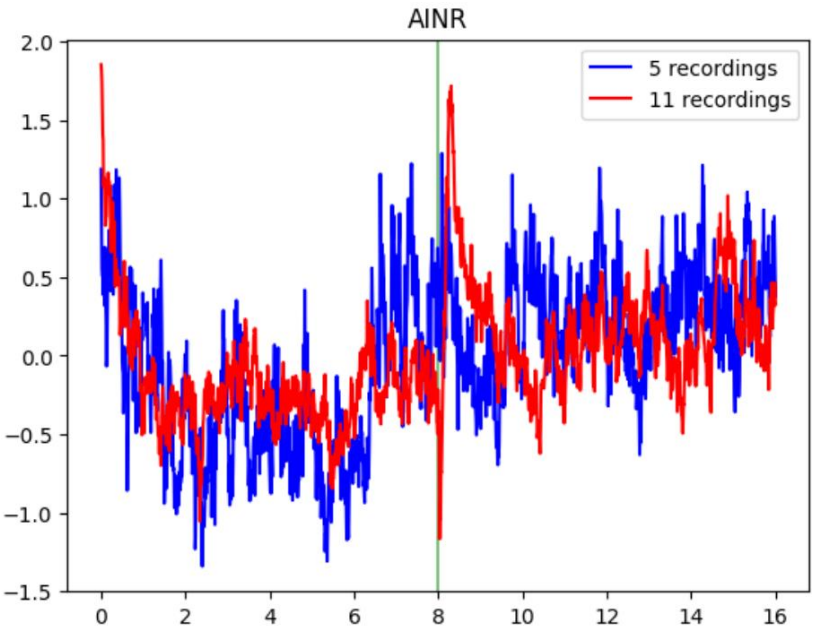
AINR:

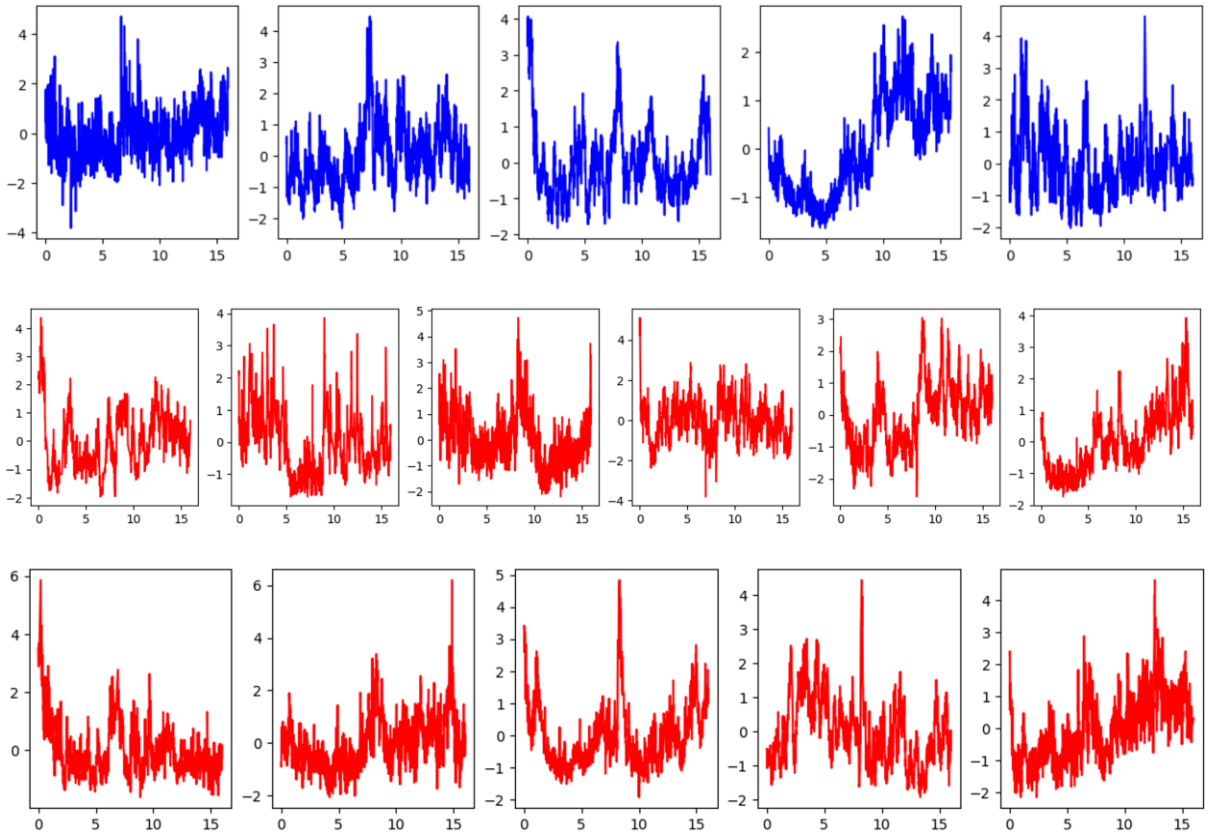
Absolute difference between avg heat and avg baseline (all time ranges normalised by dividing by the number of measurements in that time range):

- Total: 0.40580264146217304
- 6 mins - 10 mins: 0.5074397141708197
- After 8 mins: 0.42144622843959256

Average absolute difference within group (normalised):

- Baseline: 1.0640107643195122
- Heat: 1.070521634137507





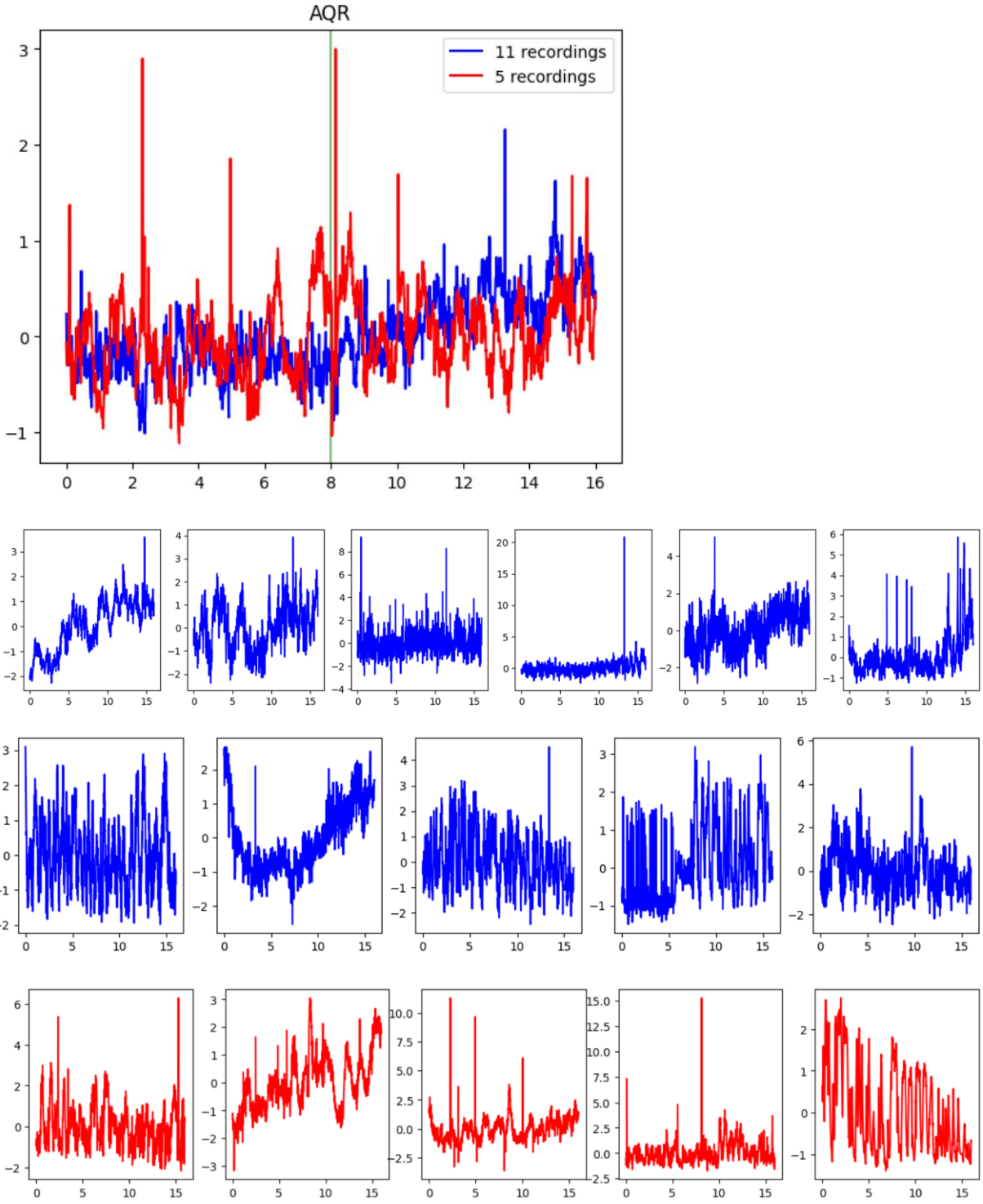
AQR:

Absolute difference between avg heat and avg baseline (all time ranges normalised by dividing by the number of measurements in that time range):

- Total: 0.4142601895657753
- 6 mins - 10 mins: 0.5073533926874011
- After 8 mins: 0.42918338430853586

Average absolute difference within group (normalised):

- Baseline: 1.0523657059218028
- Heat: 1.0941798909457625



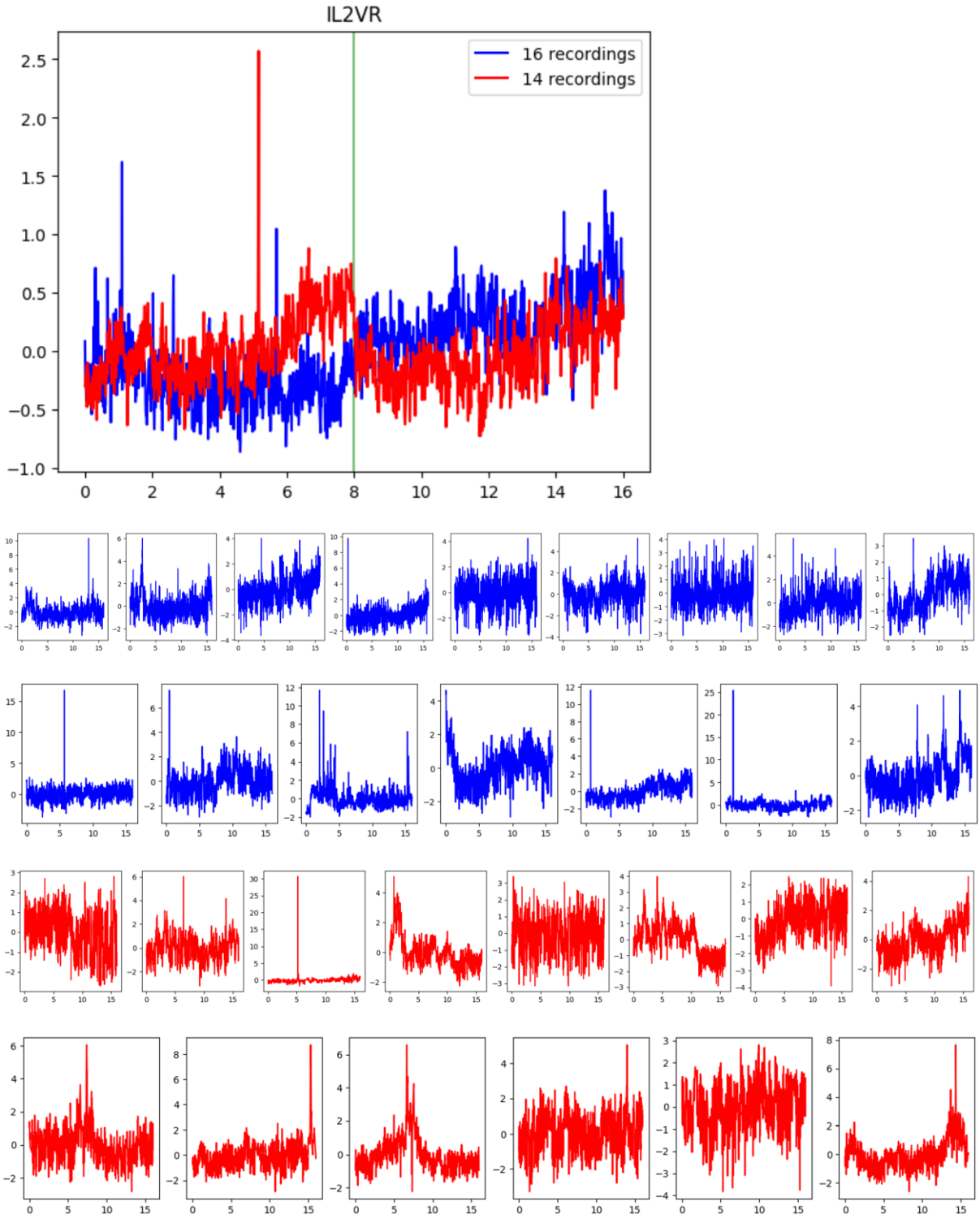
IL2VR:

Absolute difference between avg heat and avg baseline (all time ranges normalised by dividing by the number of measurements in that time range):

- Total: 0.3726270644412103
- 6 mins - 10 mins: 0.491389803575899
- After 8 mins: 0.35484468964236315

Average absolute difference within group (normalised):

- Baseline: 1.0278415844446596
- Heat: 1.06277874078395

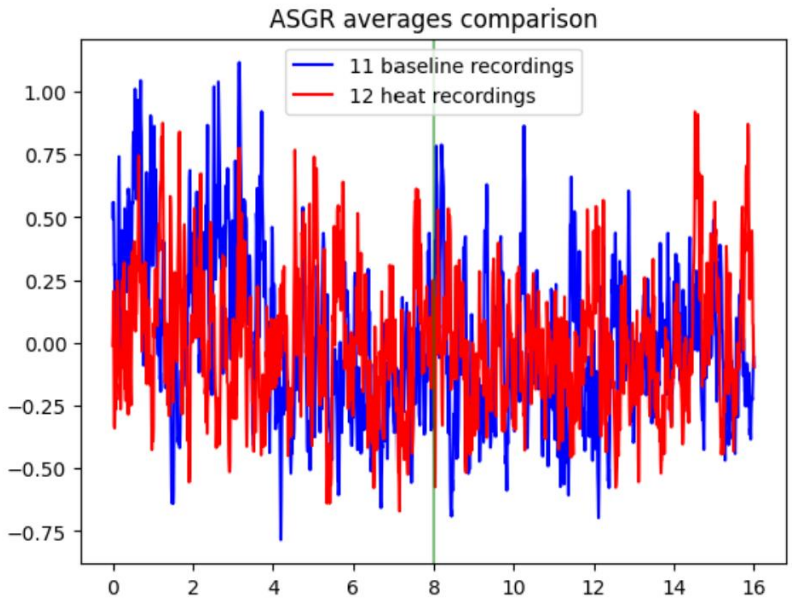


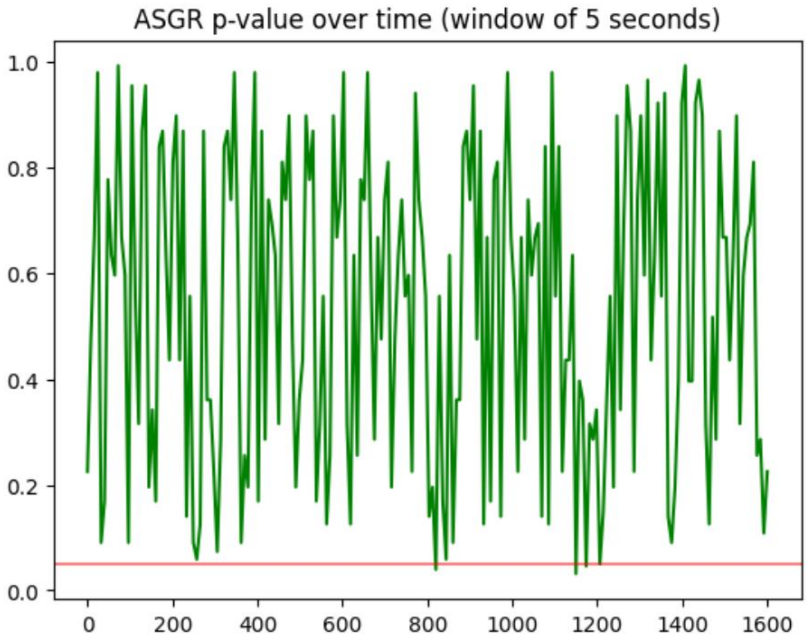
Appendix D

Graphs of p-value over time of 17 statistically significant neurons

This appendix contains the graphs of p-value over time for the 17 neurons identified as statistically significant, delay and duration values for each time window, the graphs of the averages of baseline and heat groups for these neurons, and graphs of individual measurements used to calculate the averages.

ASGR:

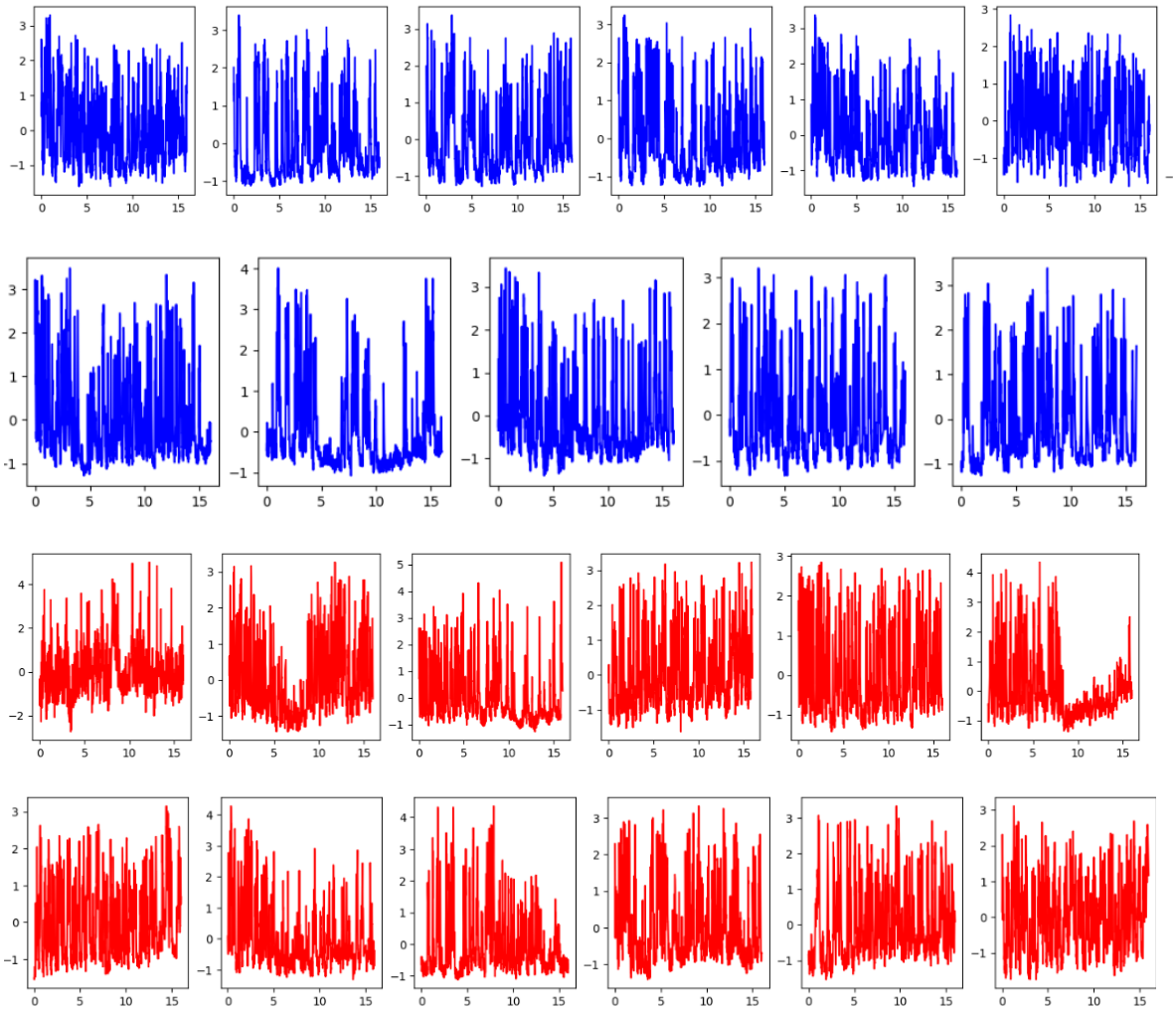




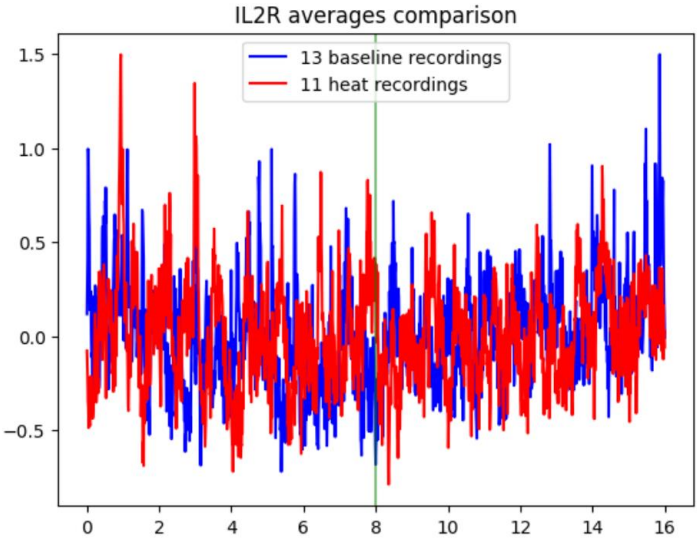
Delay: 10 seconds after stimuli

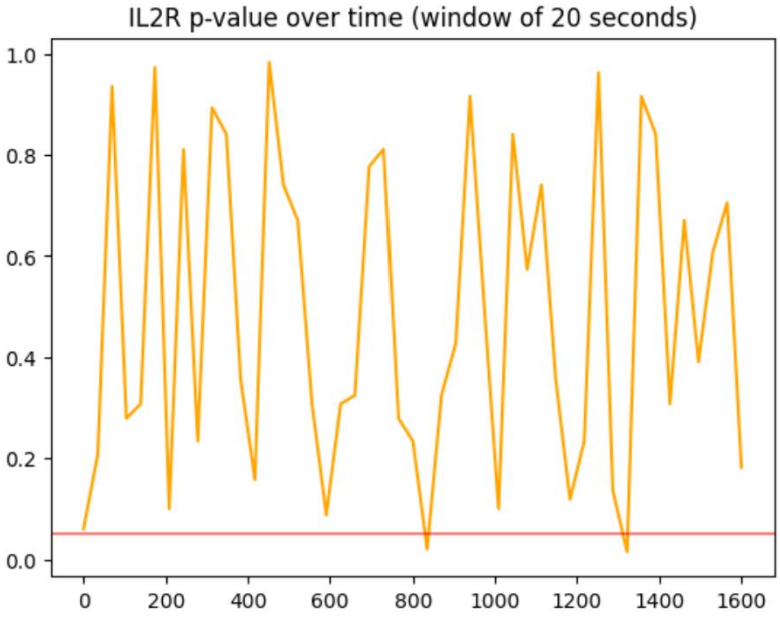
Duration: 5 seconds

Individual measurements:



IL2R:

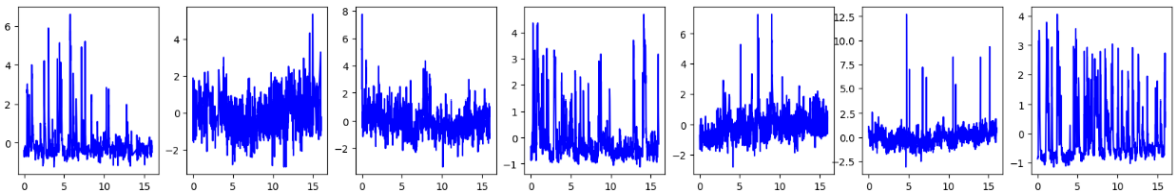


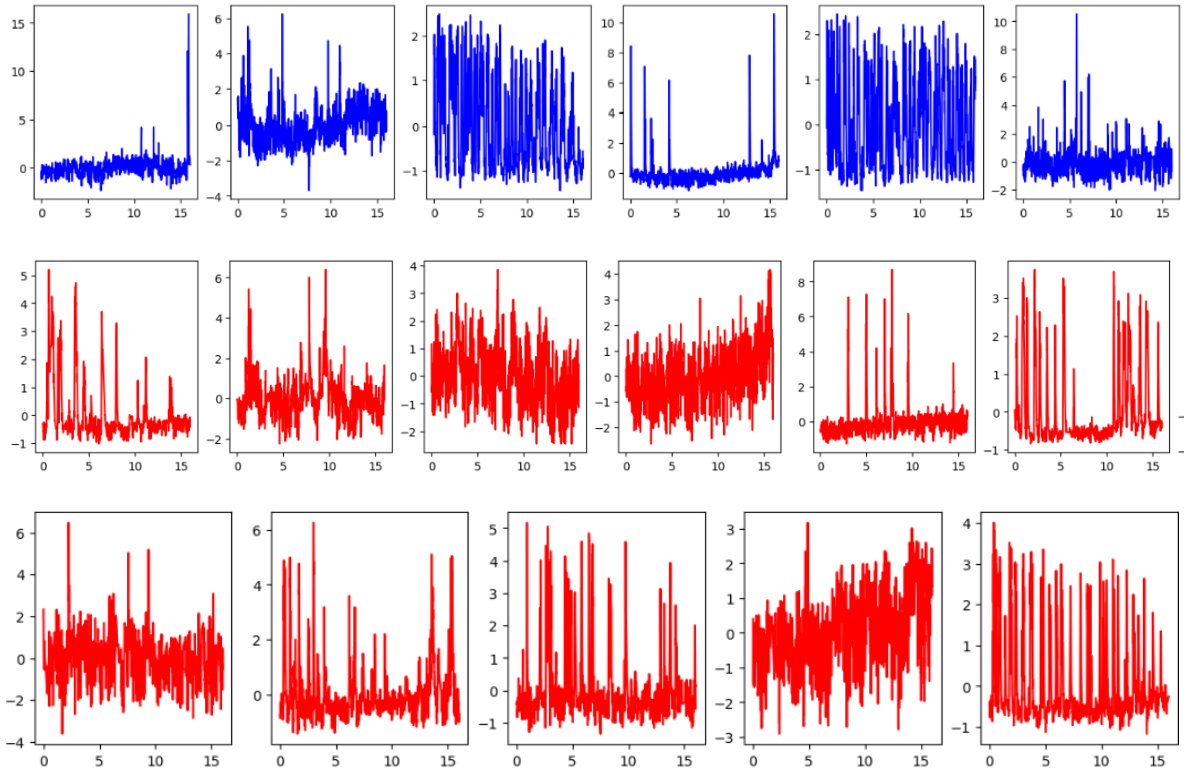


Delay: 20 seconds after stimuli

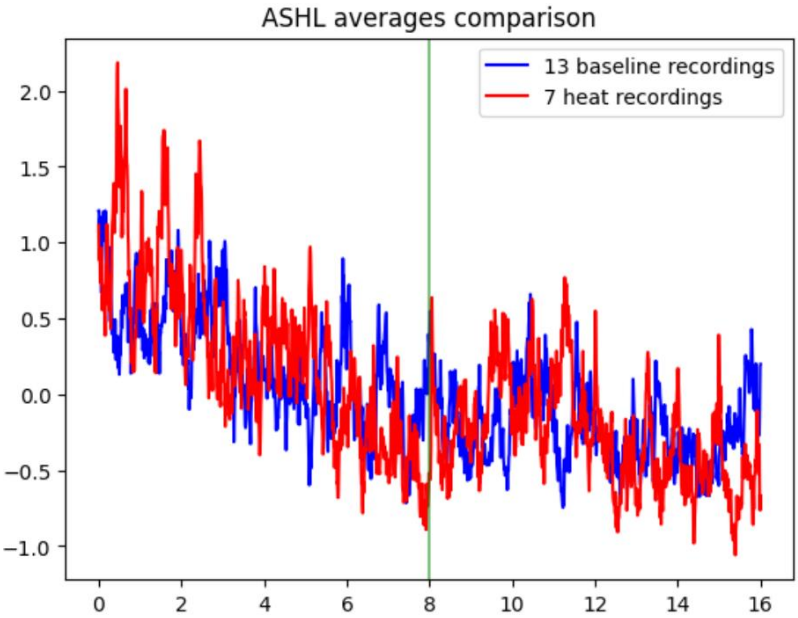
Duration: 20 seconds

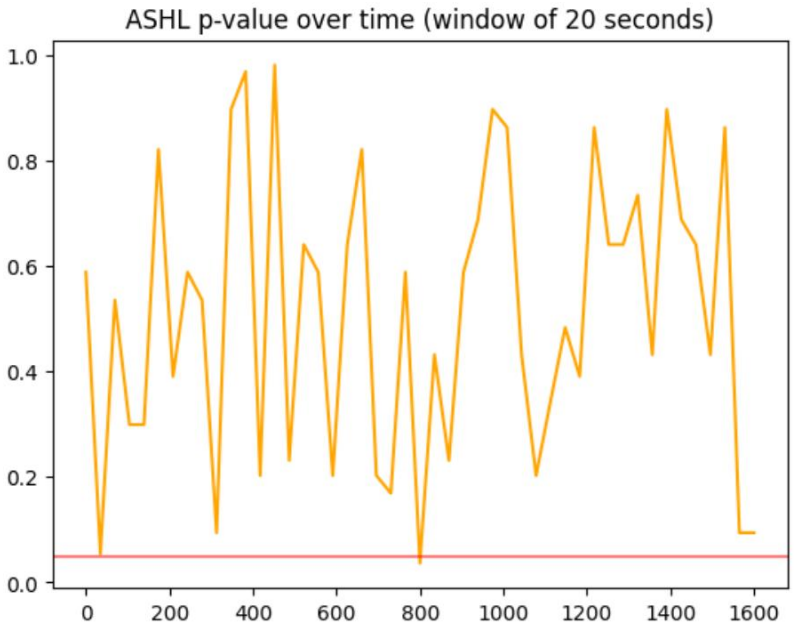
Individual measurements:





ASHL:

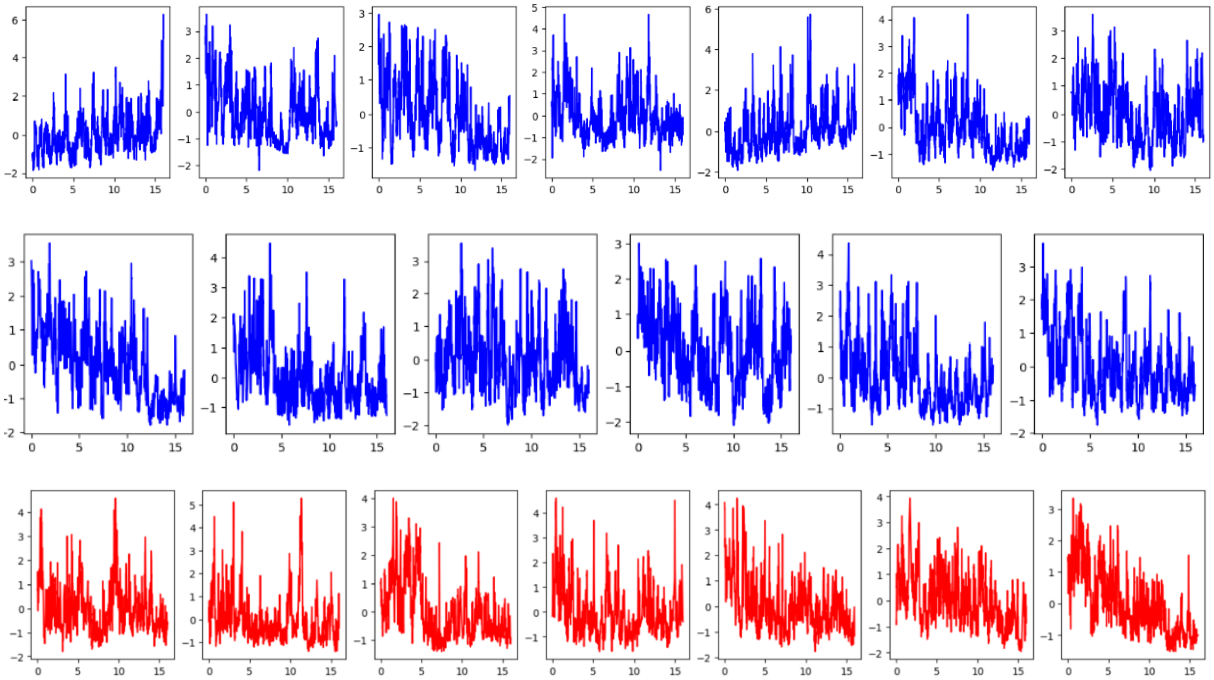

ASHL averages comparison
13 baseline recordings
7 heat recordings
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
0
2
4
6
8
10
12
14
16



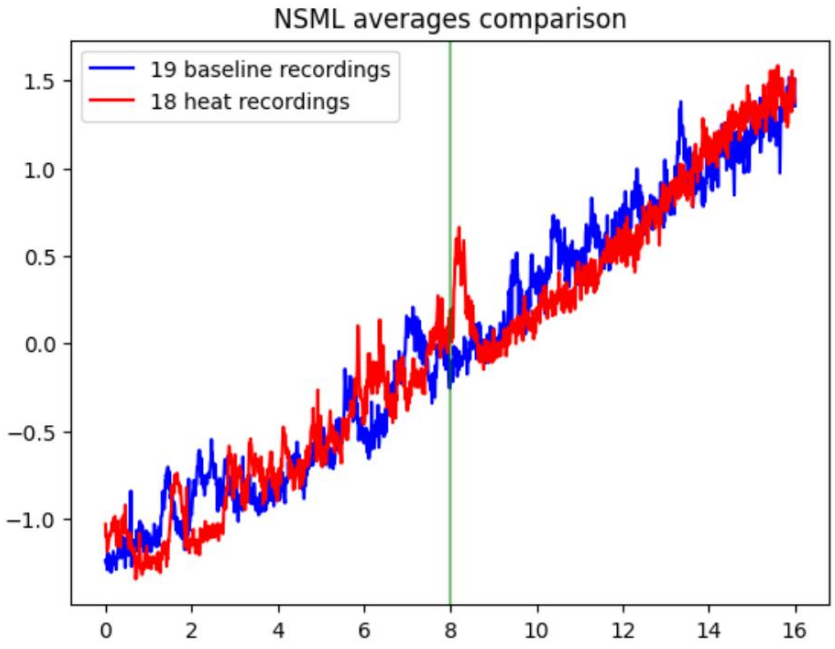
Delay: 0 seconds after stimuli

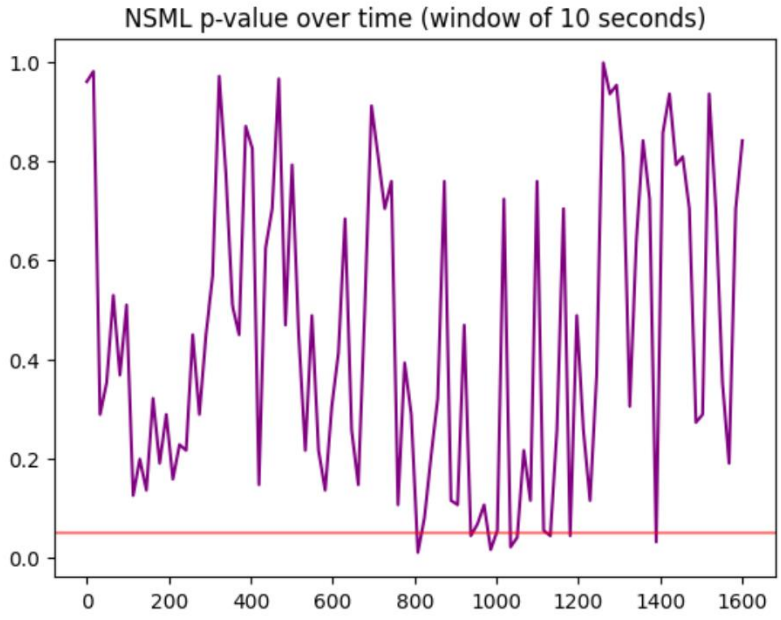
Duration: 20 seconds

Individual measurements:



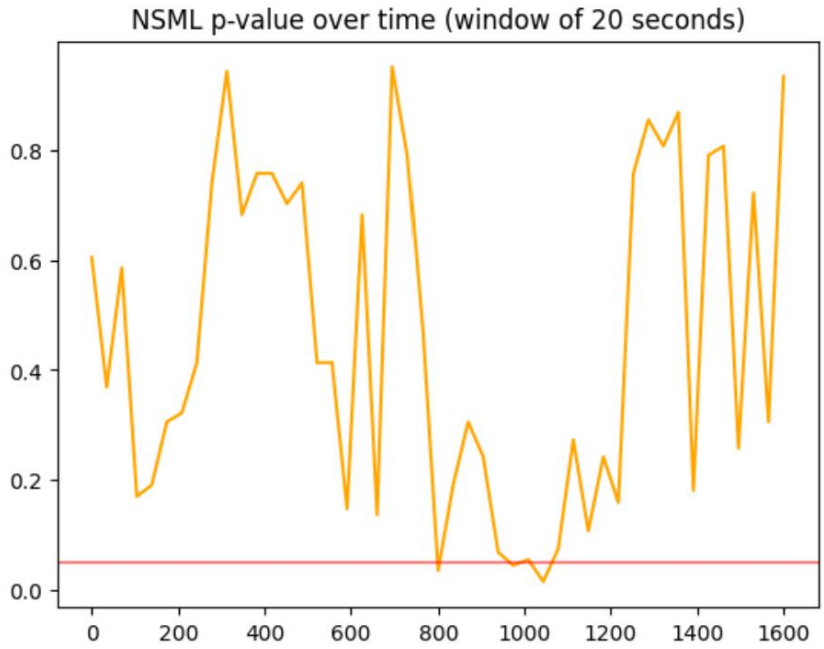
NMSL:





Delay: 0 seconds after stimuli

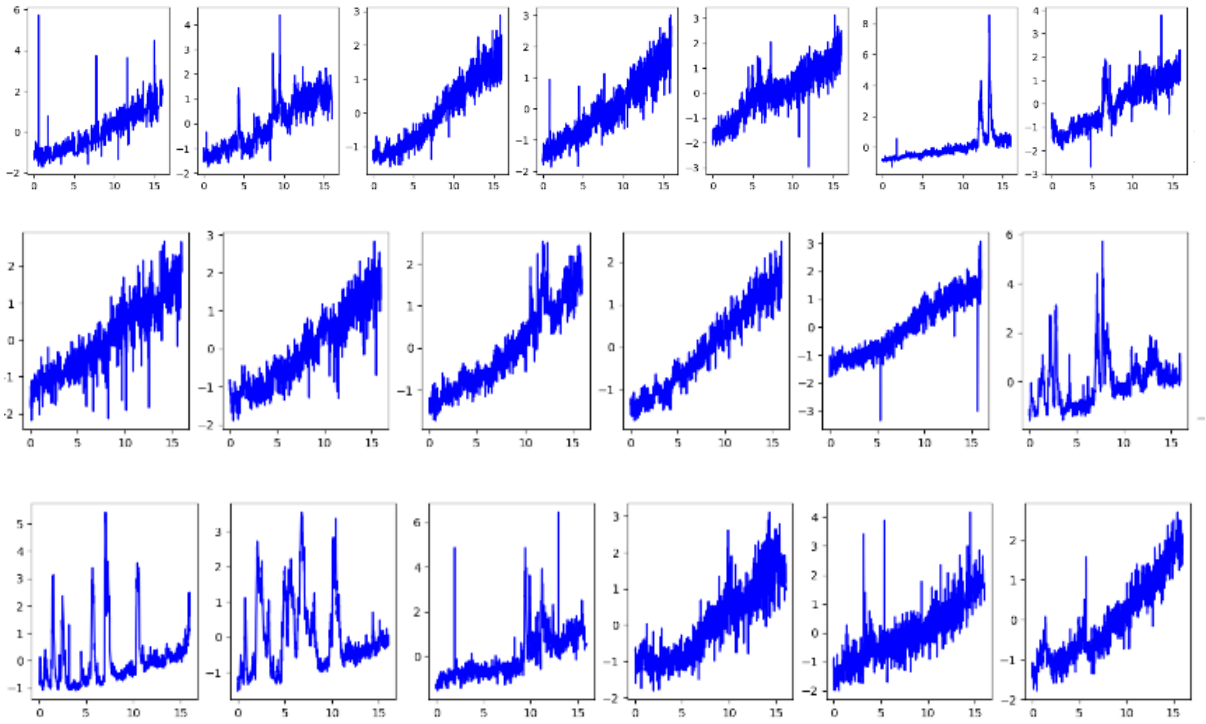
Duration: 120 seconds

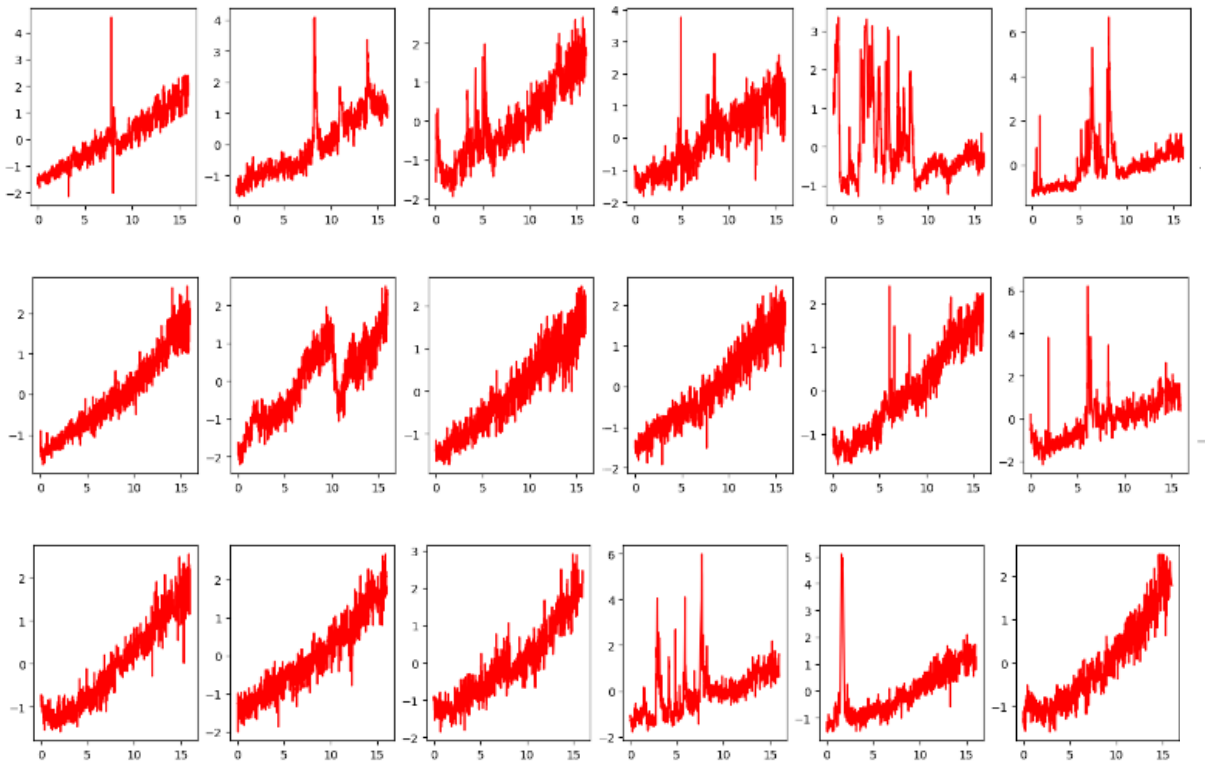


Delay: 0 seconds after stimuli

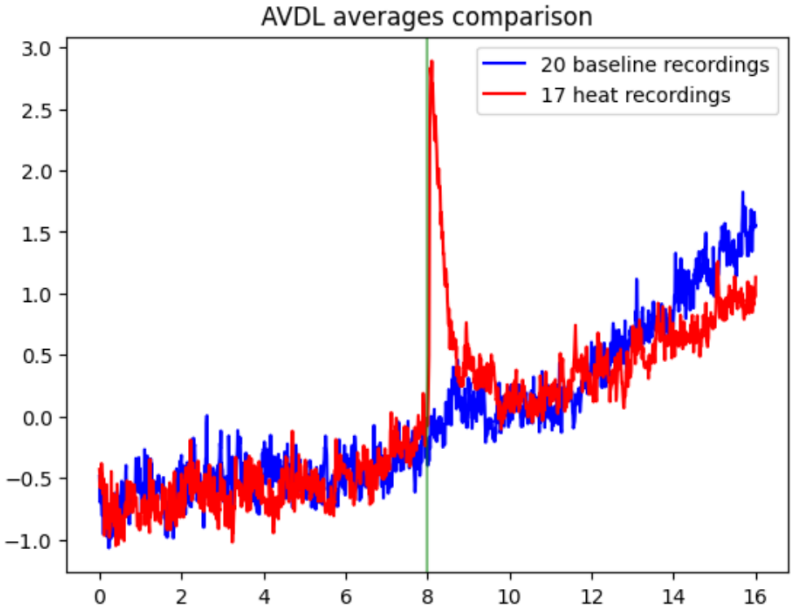
Duration: 120 seconds

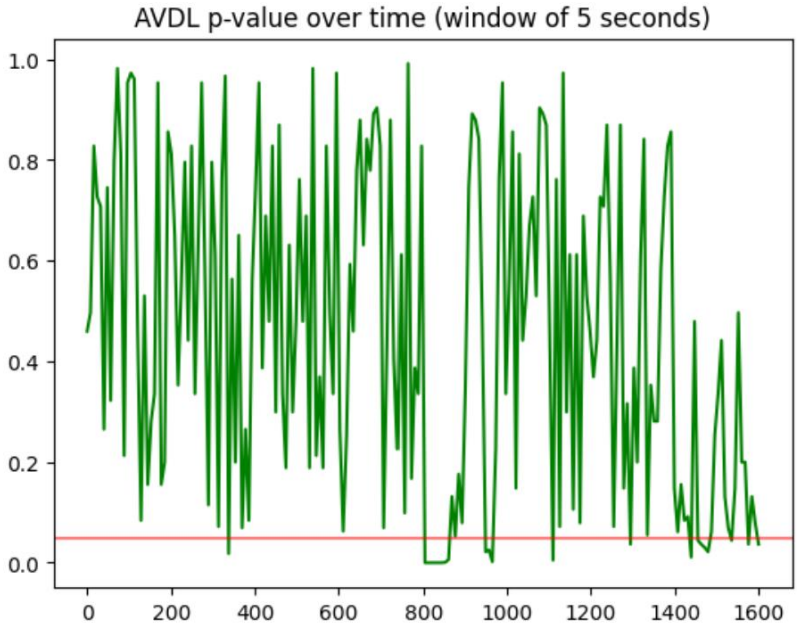
Individual measurements:





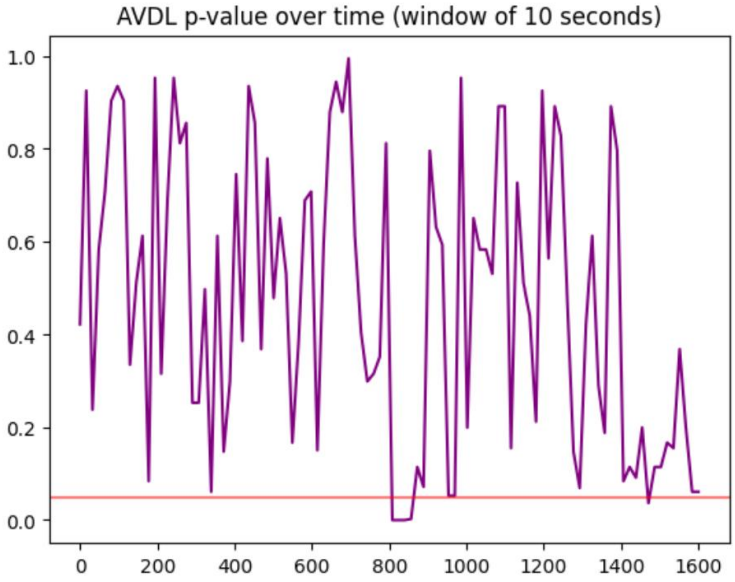
AVDL:


AVDL averages comparison
20 baseline recordings
17 heat recordings
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
0
2
4
6
8
10
12
14
16



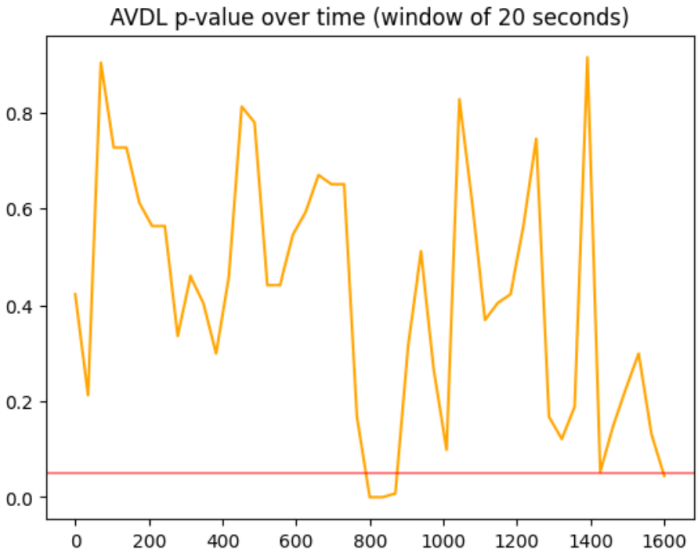
Delay: 0 seconds after stimuli

Duration: 105 seconds



Delay: 0 seconds after stimuli

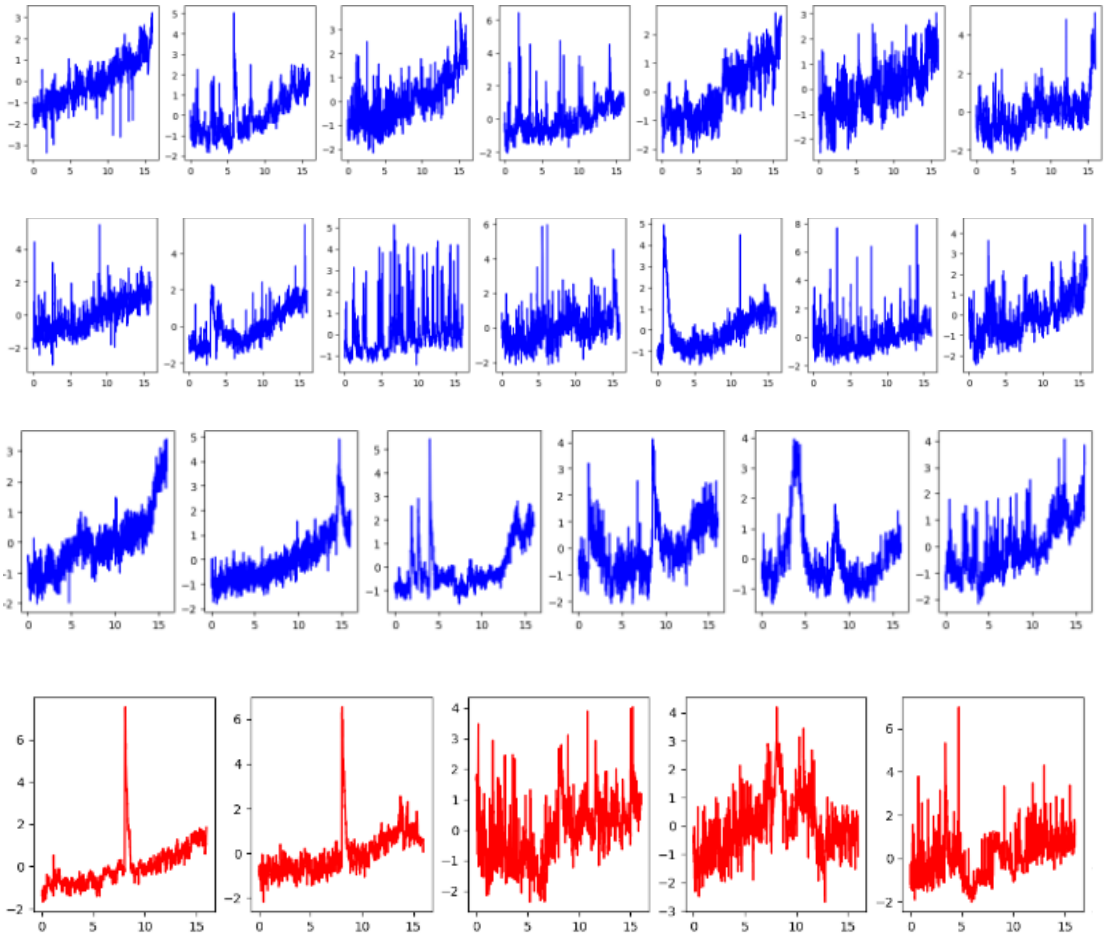
Duration: 40 seconds

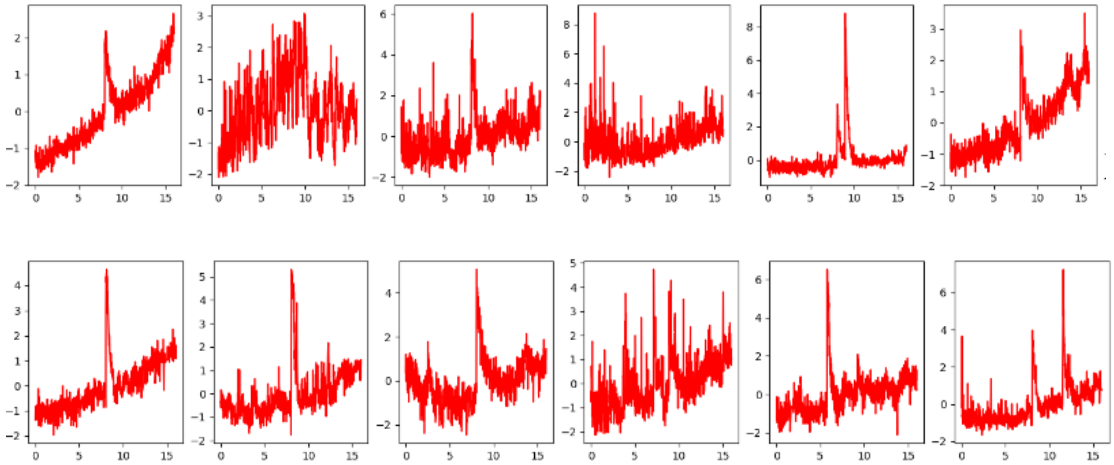


Delay: 0 seconds after stimuli

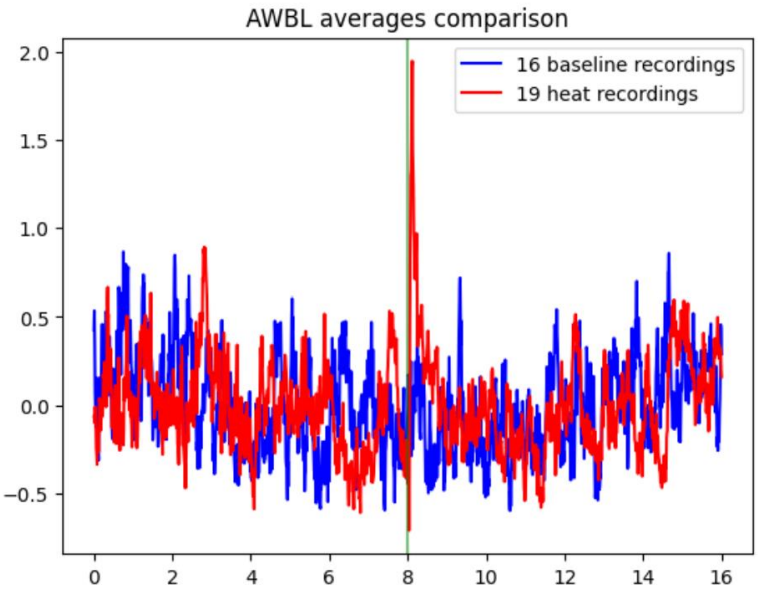
Duration: 60 seconds

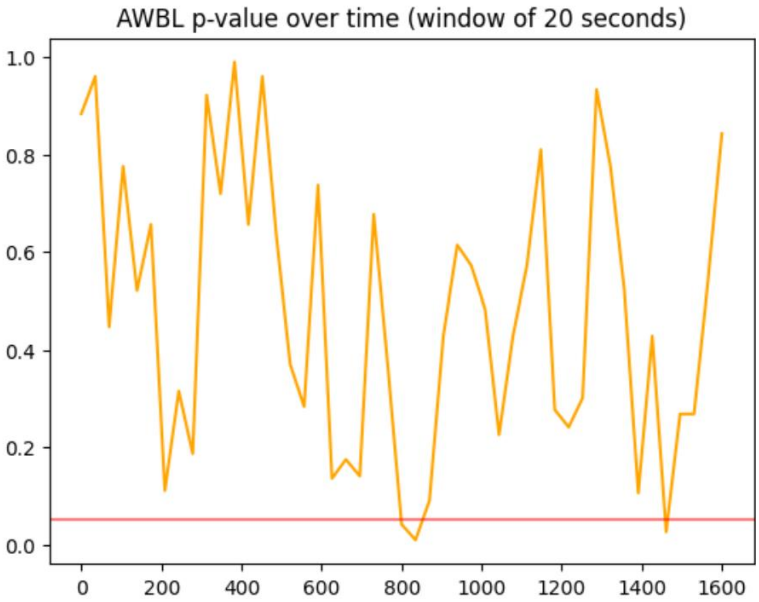
Individual measurements:





AWBL:

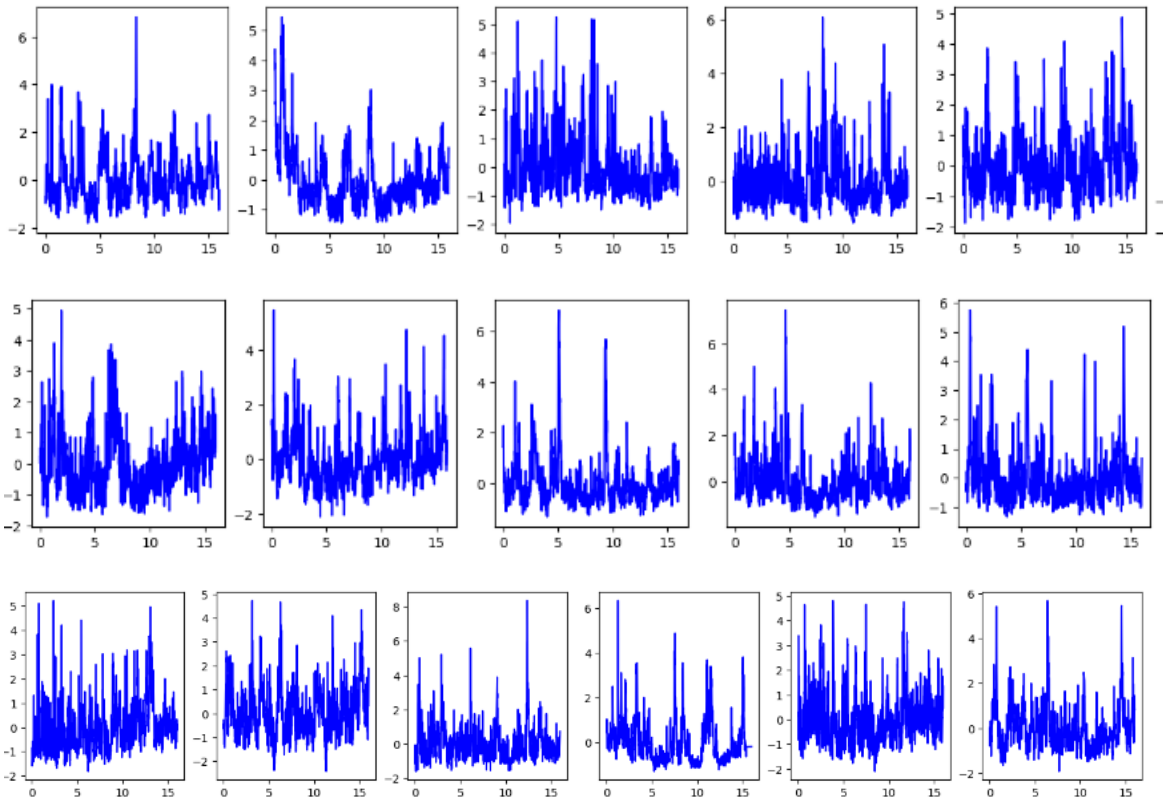

AWBL averages comparison
16 baseline recordings
19 heat recordings
2.0
1.5
1.0
0.5
0.0
−0.5
0
2
4
6
8
10
12
14
16

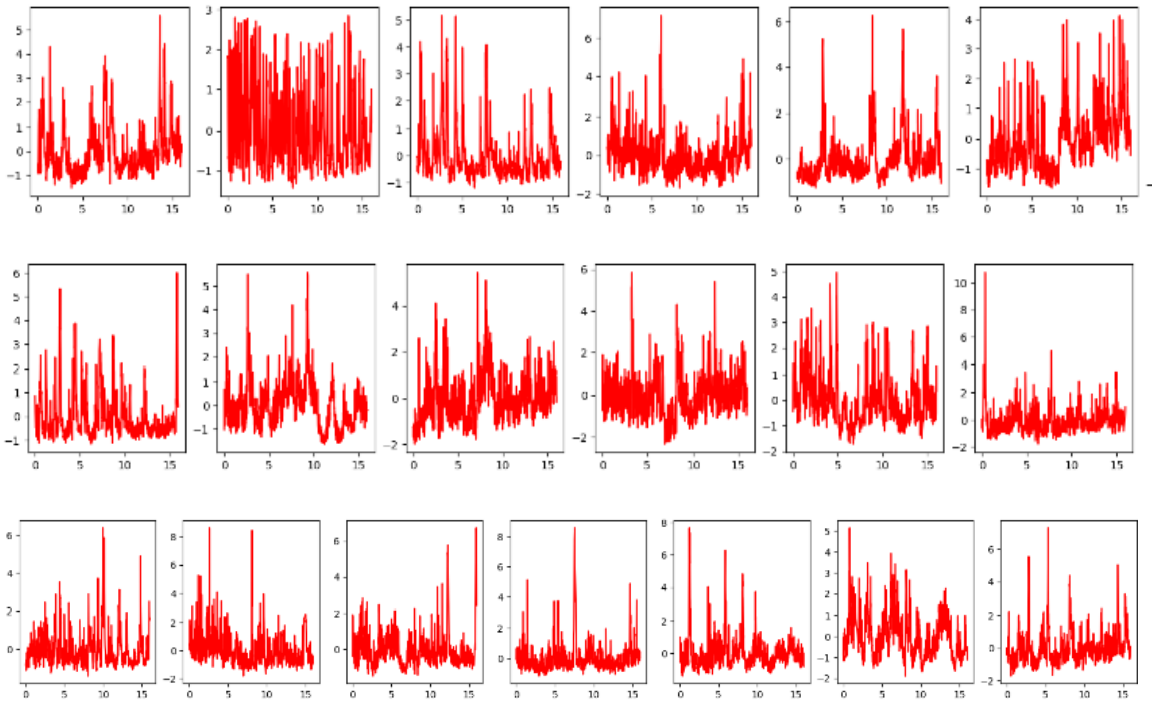


Delay: 0 seconds after stimuli

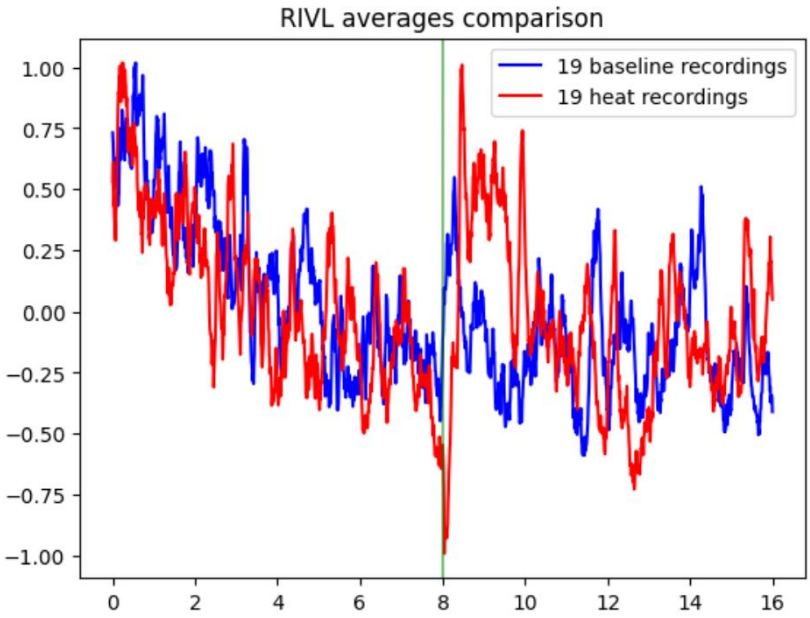
Duration: 40 seconds

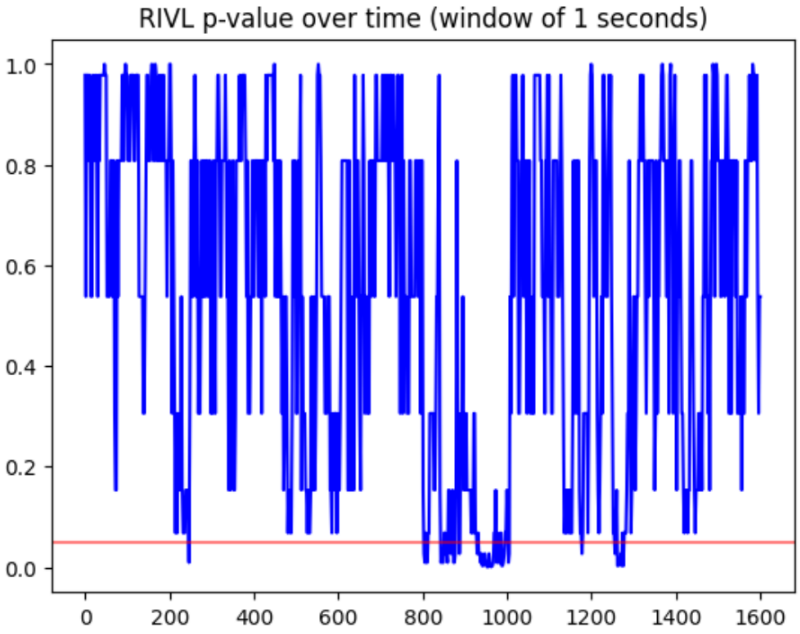
Individual measurements:





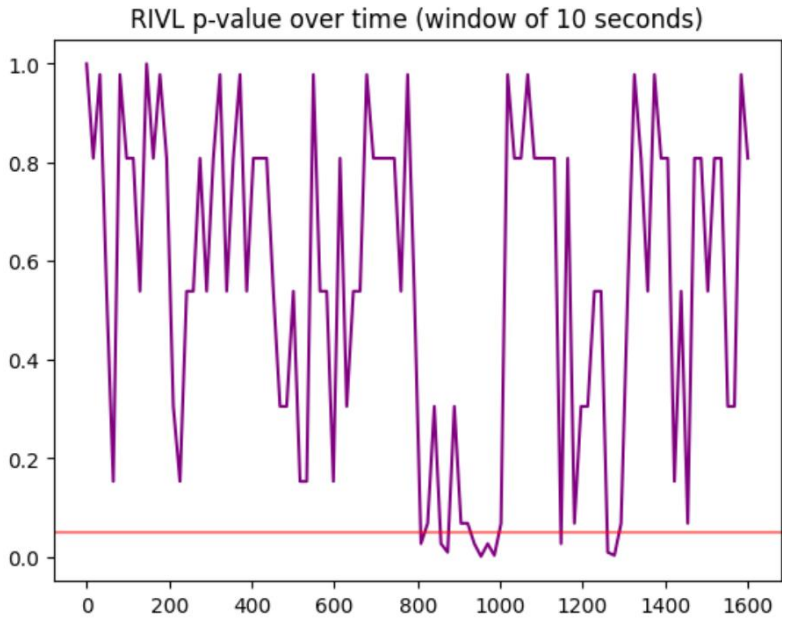
RIVL:





Delay: 1 seconds after stimuli

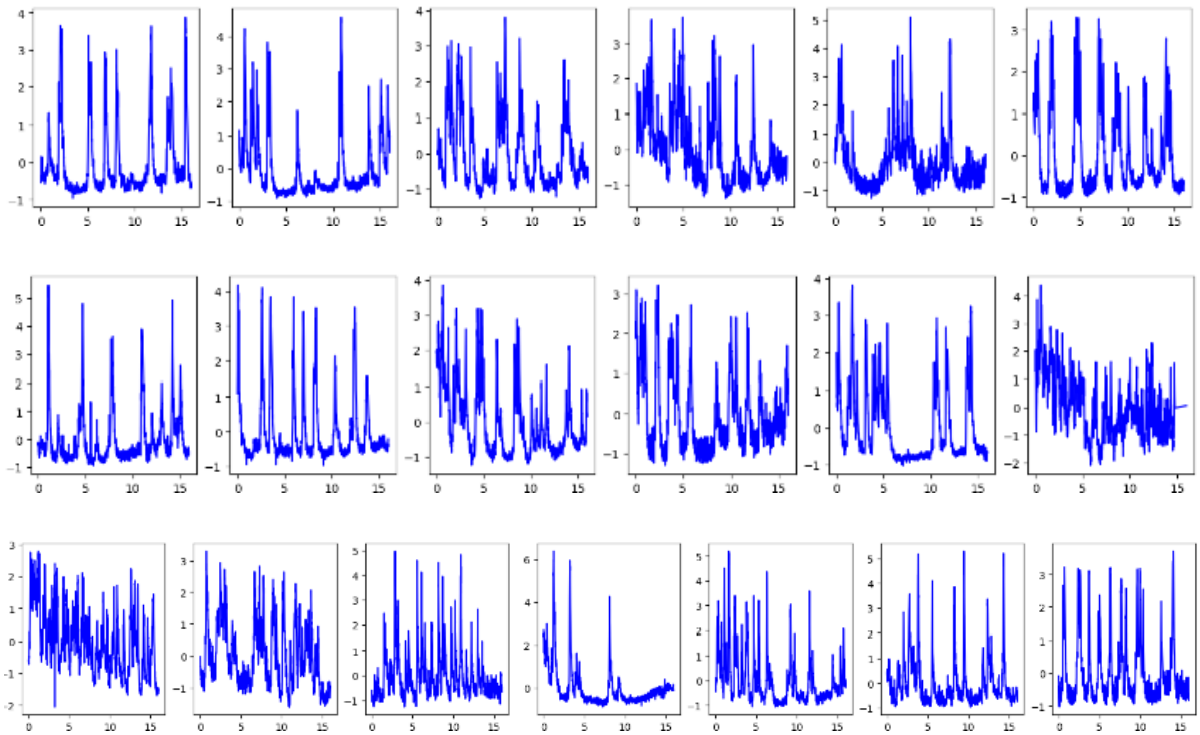
Duration: 96 seconds

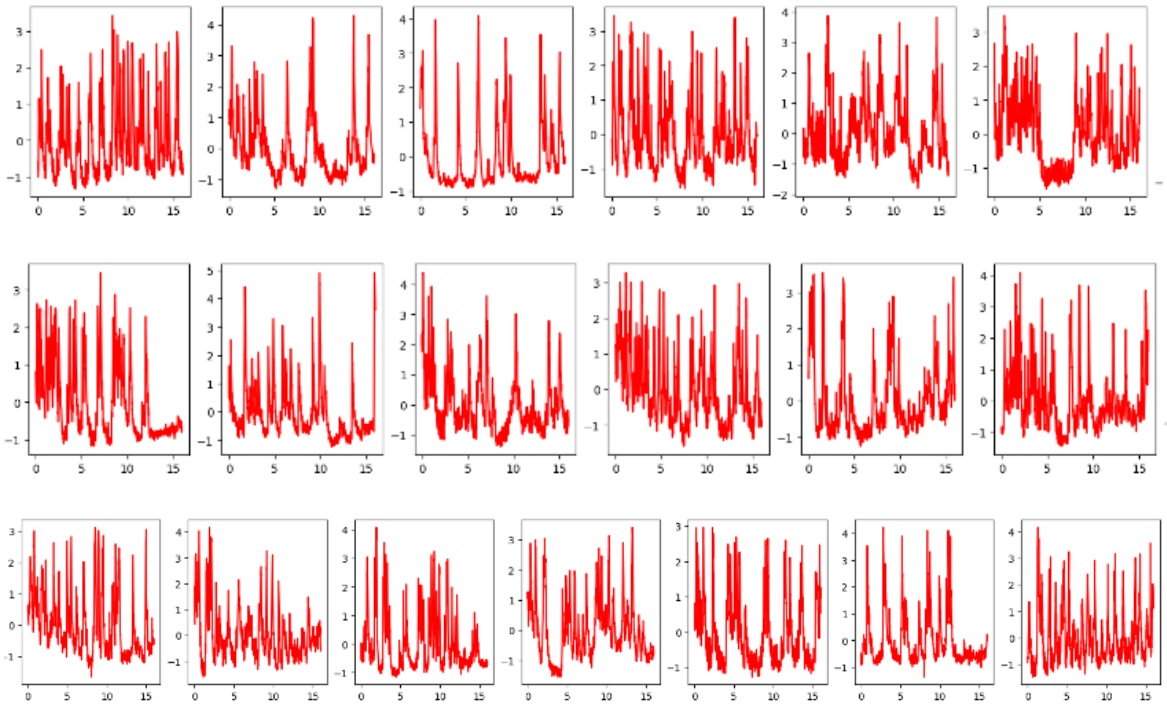


Delay: 0 seconds after stimuli

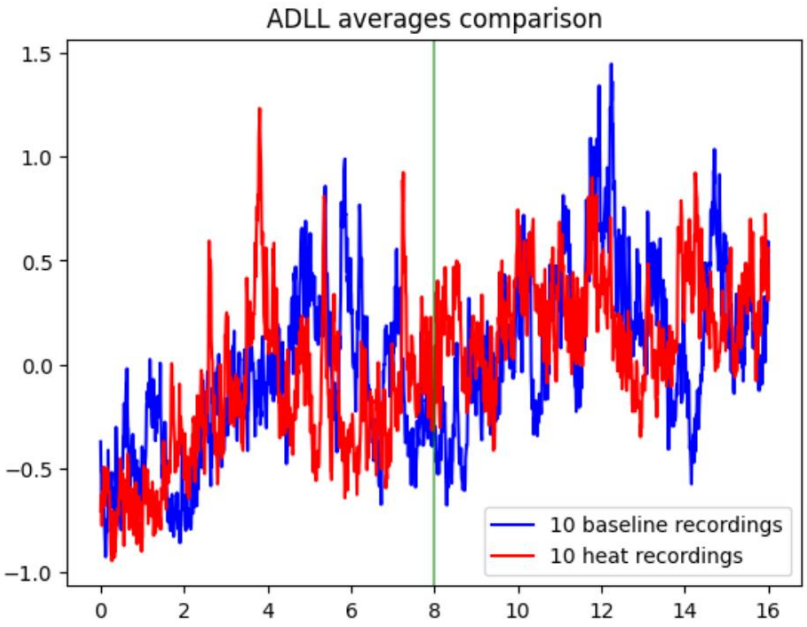
Duration: 120 seconds

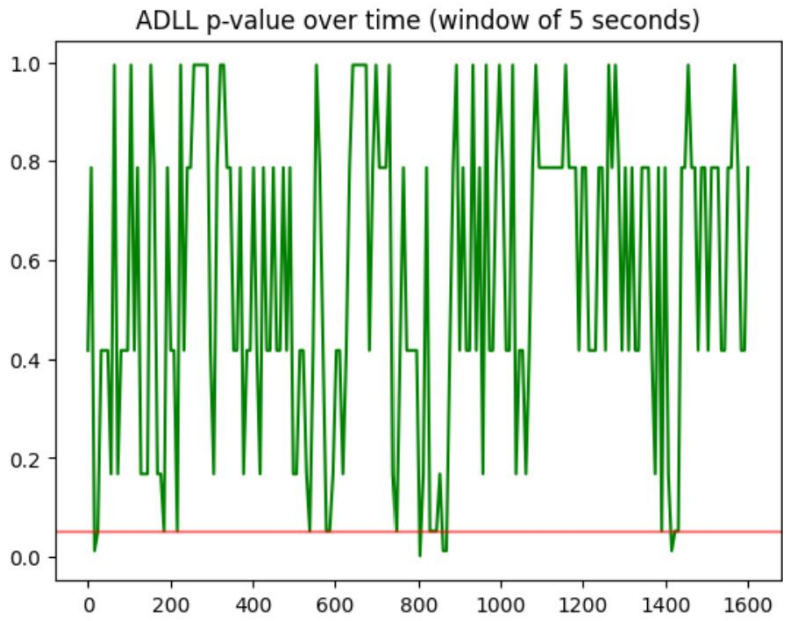
Individual measurements:





ADLL:

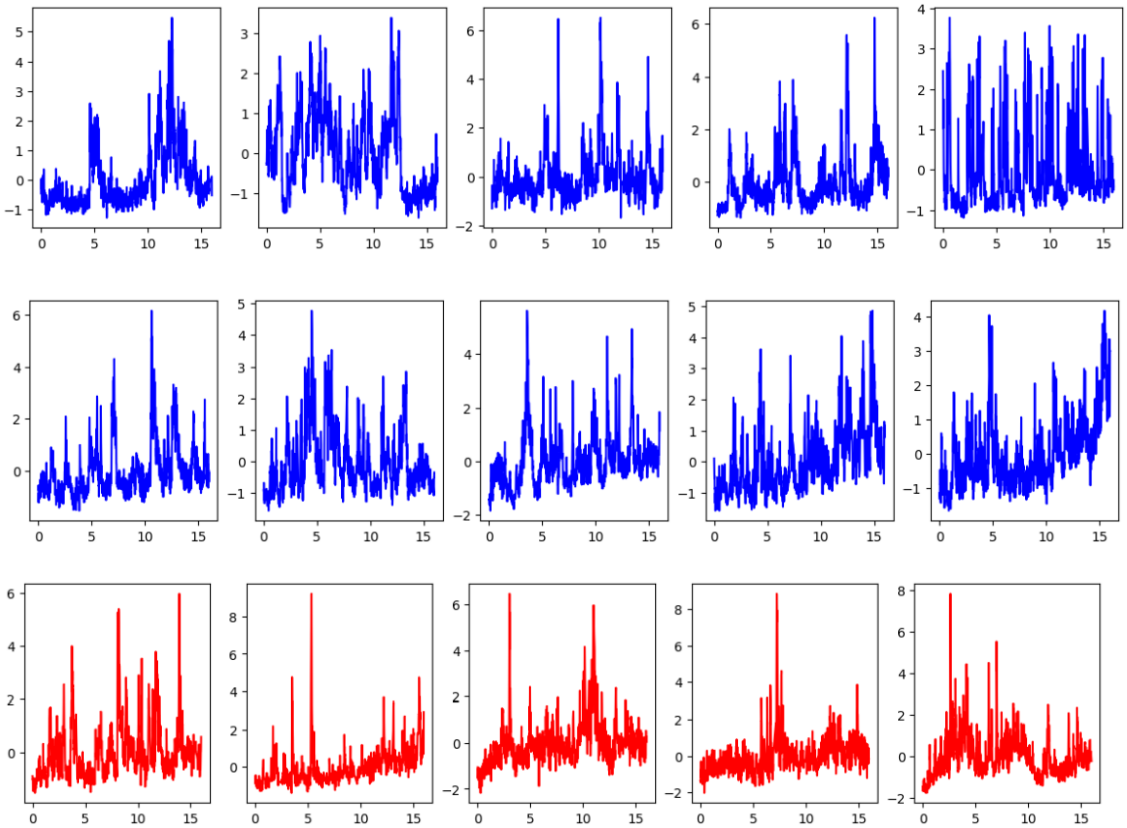

ADLL averages comparison
1.5
1.0
0.5
0.0
−0.5
−1.0
0
2
4
6
8
10
12
14
16
10 baseline recordings
10 heat recordings

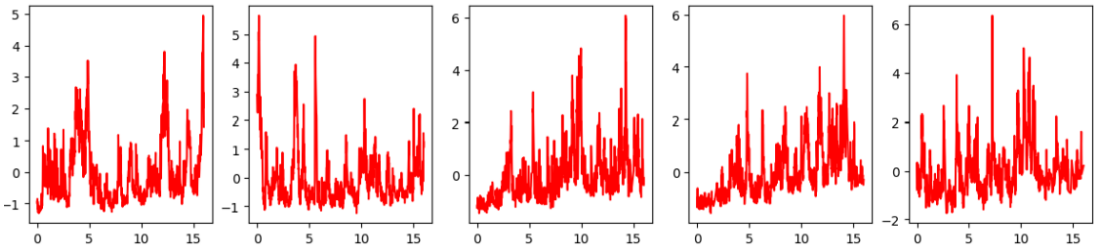


Delay: 0 seconds after stimuli

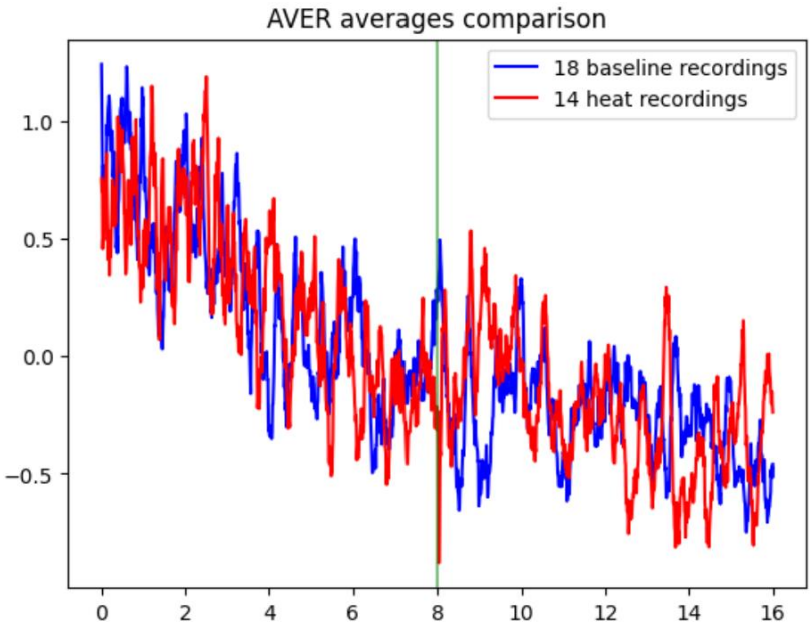
Duration: 45 seconds

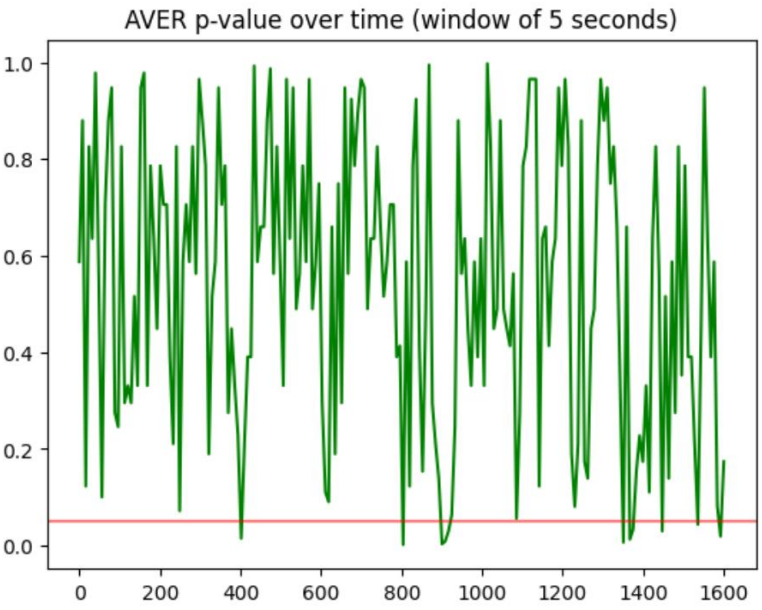
Individual measurements:





AVER:

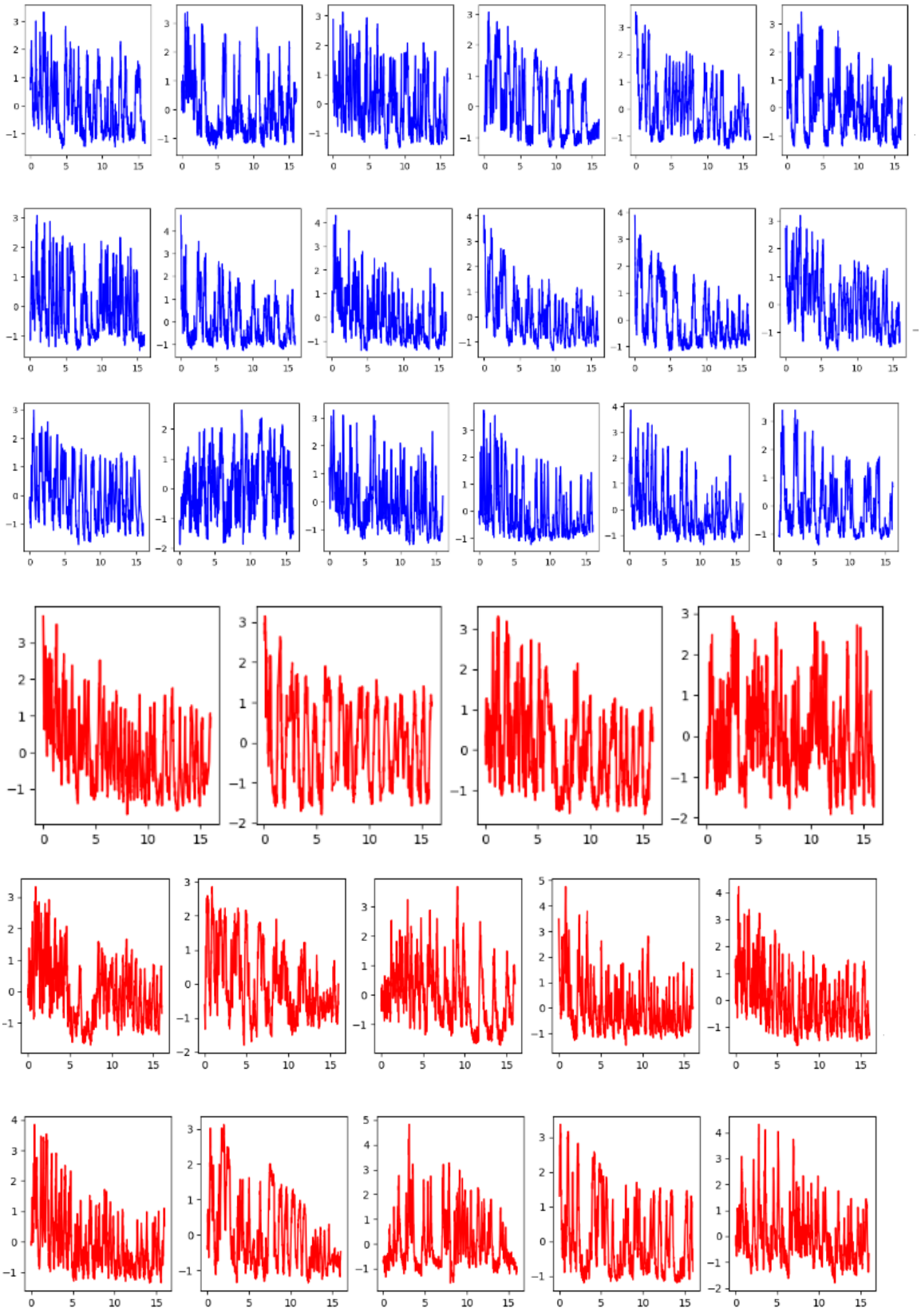




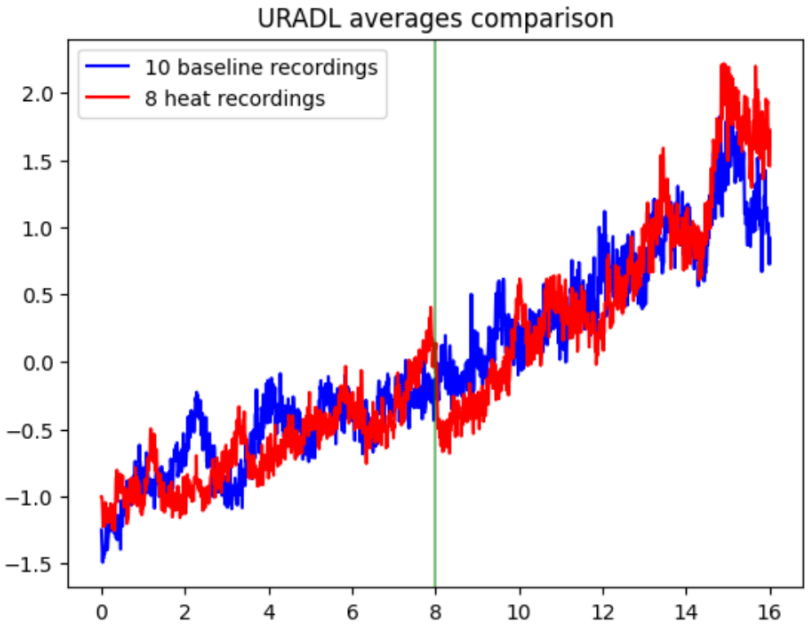
Delay: 0 seconds after stimuli

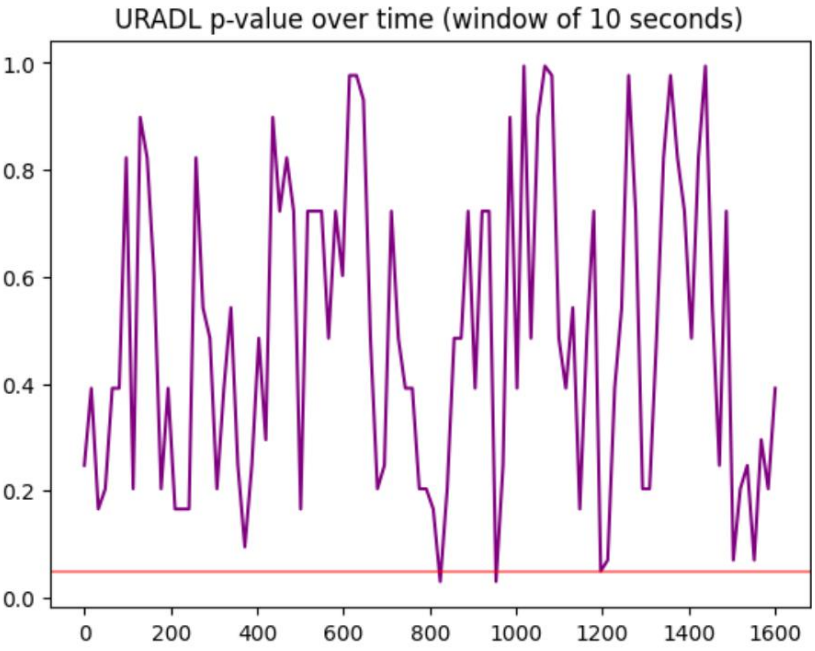
Duration: 75 seconds

Individual measurements:



URADL:

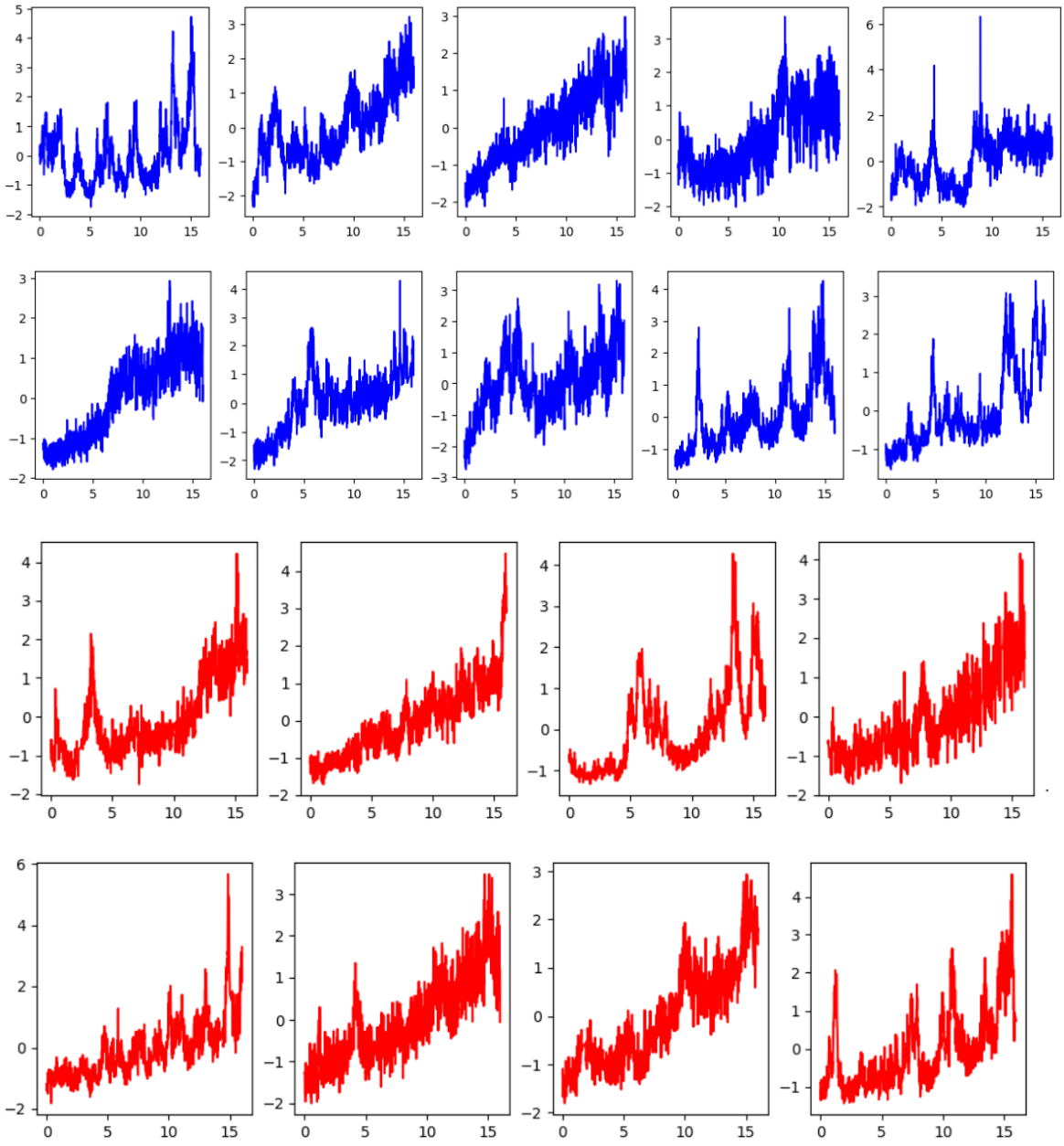




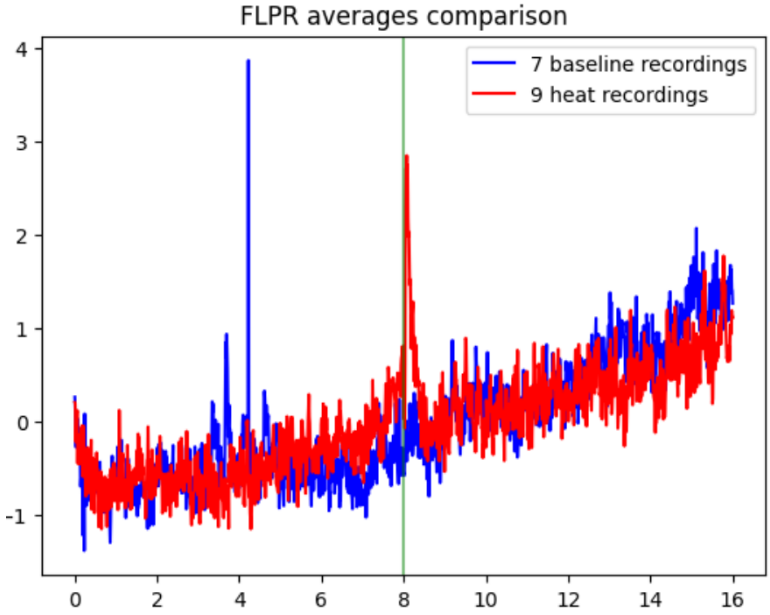
Delay: 10 seconds after stimuli

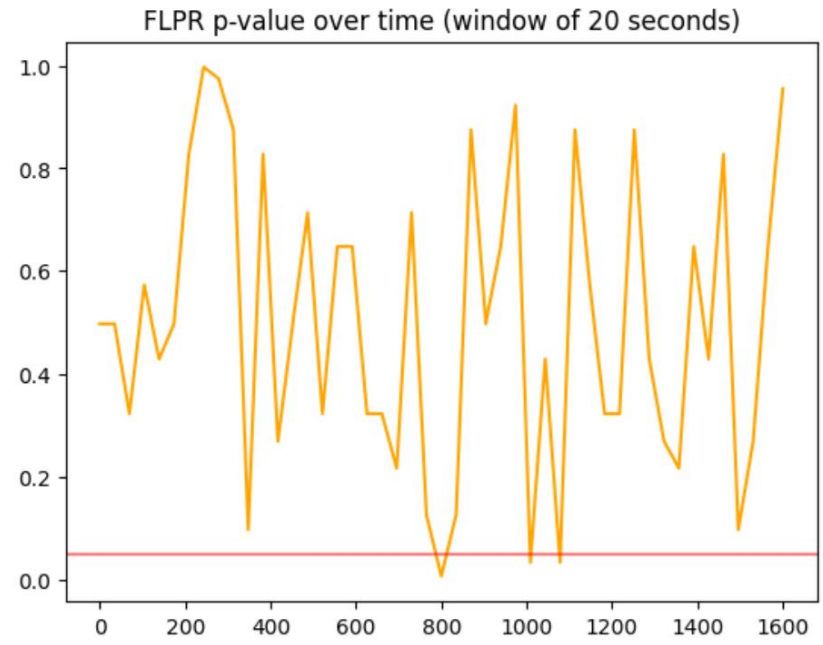
Duration: 90 seconds

Individual measurements:



FLPR:





Delay: 0 seconds after stimuli

Duration: 20 seconds

Individual measurements: