Incontri tesi IIT

F.Bernardi - G.Fregiari - R.Sacco - G.Chiaravalli

October 28, 2021

1 Incontro 1 (05/10/2021)

1.1 Spiegazione dell'esperimento

L'esperimento di interbrain svolto sui topi si pone l'obiettivo di studiare l'attività cerebrale (attività del calcio intracellulare) in 3 tipi di cavie: **observer, neutral, altered**.
Il topo observer è libero di muoversi in un'arena in cui, in due gabbie distinte, sono presenti anche i topi neutrale e alterato. Il topo è alterato se il suo *affective state* è modificato in negativo (mancanza di acqua, topo stressed), o in positivo (dopo mancanza di acqua, acqua a volontà).
Gli obiettivi dell'esperimento possono essere molteplici: monitorare l'attività calcio nell'observer e vedere come cambia quando esso si trova in prossimità di neutrale e alterato, trovare pattern nell'accensione dei neuroni ecc... Si vuole inoltre studiare la *sincronizzazione* neuronale: attività calcio simili tra due topi che interagiscono, ed eventuale imposizione del ritmo da parte di uno dei due (parte di interbrain, vedi articolo Kingsbury).
L'esperimento consiste di 3 fasi:

- **homecage**: i topi sono in una gabbia, fermi, isolati dagli altri. Per 5 minuti si monitora l'attività calcio
- **habituation**: il topo observer è libero di girare per l'arena, ma senza poter iteragire con gli altri due. Si vuole vedere come è, in condizioni isolate, l'attività dell'observer e il suo movimento nell'arena, per 15 minuti
- **test**: il test vero e proprio con tutti e 3 i topi, per 15 minuti

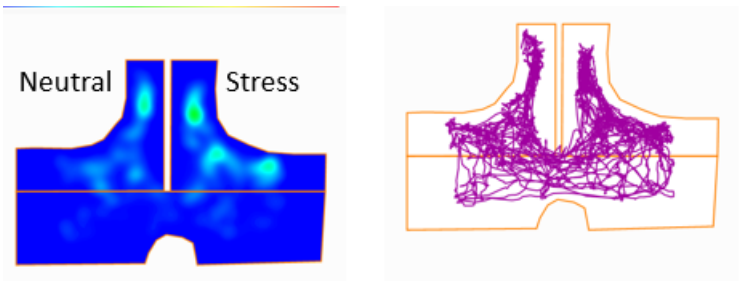

Figure 1: A sinistra l'arena con le due gabbie in corrispondenza dei nomi neutral e stressed. A destra il movimento effettuato dall'observer durante il test

1.2 Inscopix e file excel

Inscopix è l'azienda che produce i microscopi che inseriti nel topo, ne monitorano l'attività. Prima del test, viene iniettato nel topo un virus che reagisce con il calcio provocando fluorescenza. Tale fluorescenza viene misurata per ogni frame come $\frac{\Delta F}{F_b} = \frac{F(x,y,t)-F_b}{F_b}$, dove $F(x,y,t)$ è la fluorescenza misurata nel punto (x,y) al tempo t, mentre F_b è la fluorescenza baseline, scelta in genere come la media o la minima (vedi pag. 41 manuale). Dai video ottenuti lungo le 3 fasi dell'esperimento, attraverso diversi algoritmi di identifiifcazione, filtering ecc (vedi manuale), si producono diversi file excel, da analizzare su Matlab:

- Un file per ognuno dei 3 topi con i tempi nella prima colonna e, nelle successive, ogni colonna per un neurone, che alla riga corrispondente a un tempo avrà il valore misurato $\frac{\Delta F}{F_b}$. La prima riga dice se il neurone è preferibile da accettare o rifiutare a seguito del processing di inscopix
- Due file tengono traccia della posizione dell'observer (uno durante l'habituation, uno durante il test), indicando le sue coordinate e la sua vicinanza o meno alle gabbie degli altri due topi con valori binari
- Un file contiene informazione sull'attività (sniffing) dell'observer in interazione con gli altri topi

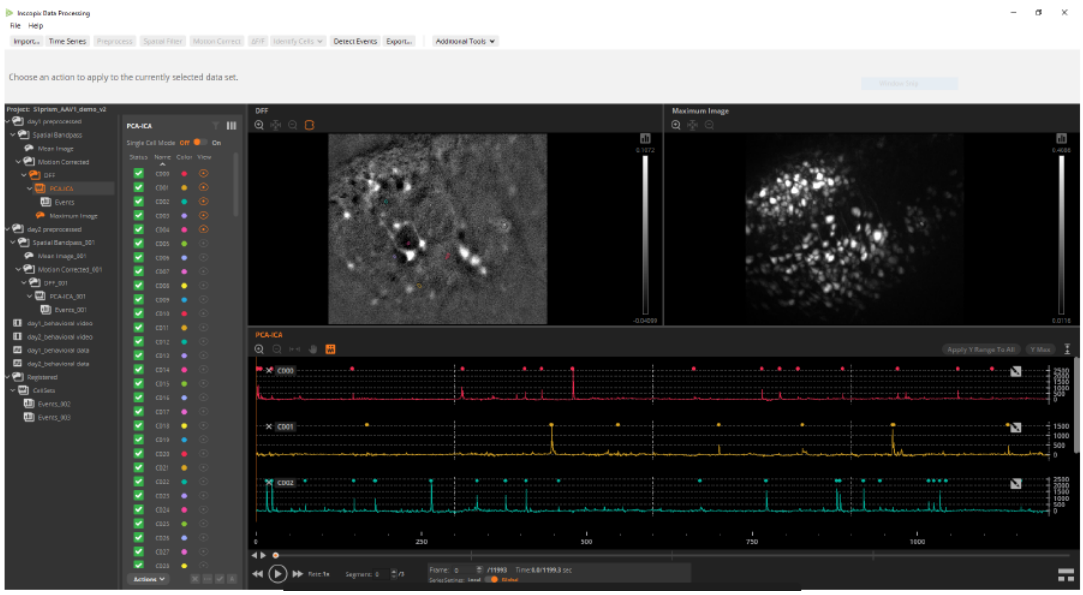

Figure 2: Esempio video processato da Inscopix

1.3 Discussione

L'analisi del dataset può percorrere tantissime strade diverse. Si è per esempio discusso di:

- Il metodo più adeguato per rappresentare l'attività del calcio estesa al singolo topo, anzichè a un singolo neurone. Una media semplice è sufficiente? Idea di considerare una media pesata per ogni neurone, dove i pesi vengono determinati in base a quanto il neurone è correlato all'interazione del topo con gli altri due (esempio, modello lineare generalizzato con attività dei neuroni come covariate, risposta come attività di sniffing verso un topo). così facendo si vede l'attività media del topo intesa come attività dei neuroni direttamente coinvolti nell' interazione (e magari sincronizzazione) (Non so se sta roba che ho scritto ha senso LOL)

- Riccardo suggerisce che, avendo l'attività calcio dell'observer, possiamo pensarla come funzione che varia nello spazio, durante il test in cui l'observer si muove nell'arena. Così facendo possiamo per esempio plottarla nel dominio dato dall'arena stessa ed associare ad ogni punto dell'arena una concentrazione calcio, per poi vedere cosa succede nelle aree vicine ai topi neutrale e stressato

- In parallelo all'analisi dei dati, si può lavorare alla **modellistica del calcio**. Problema: tali modelli studiano i meccanismi intracellulari di bilancio del calcio, ma il microscopio non ha una risoluzione così elevata da fornire validazioni sperimentali. Greta C. suggerisce di pensare a ordini di grandezza più elevati, per esempio alla diffusione del calcio tra i neuroni della rete osservata.
  Non discusso ma lo dico ora: Fabrizio si chiede se è possibile comunque pensare a un modello a "scatola chiusa", in cui ciò che succede all'interno del neurone lo si cerca di implementare da ciò che si è fatto in letteratura, ma in cui si personalizza al caso nostro focalizzandosi su ciò che il modello "scatola chiusa" prende dall'esterno per caratterizzare ogni singola simulazione dell'onda calcio (TBC).

Per quanto riguarda l'analisi dei dati, i primi passi saranno:

- funzioni Matlab che scompongono il segnale nelle 3 fasi dell'esperimento dai file excel originali

- **normalizzazione del segnale**, per esempio *normalization z-score* $z = \frac{x-\mu}{\sigma}$ o *min-max normalization* $x' = \frac{x-min(x)}{max(x)-min(x)}$

- **Statistica descrittiva** sul campione di dati: giù di medie, variabilità, boxplot e istogrammi

- **identificazione dei picchi** : quando possiamo considerare un neurone attivo? Una soluzione può essere considerare tutto ciò sopra a un certo valore come attività e tutto ciò sotto no, utilizzando come treshold, ad esempio, $y = \mu + \sigma$, media del segnale + deviazione standard (tutto ciò che è sopra tale retta può andare a costituire un 1 in un vettore di 0/1 che ci dice per ogni neurone se esso è attivo)

- **Correlazioni tra neuroni**: c'è qualcosa dalla letteraura (vedi parte di metodi nell'articolo Scheggia - Manago), ma inizierei con un'analisi standard (matrice covarianza e correlazione di Pearson)
- Differenze tra attività medie in due topi a contatto (errore L^2 ecc...)

2 Incontro 2 (13/10/2021)

2.1 What has been done

I started to write some matlab scripts using the excel data coming from Inscopix (the first 2 points include work done after the meeting, when I solved the time-adapting problem we discussed):

- The function *accept_and_split* ,which takes the raw files as input, excludes the neuron marked as *rejected* from Inscopix. It then splits the data for every mouse into the 3 parts corresponding to homecage, habituation and test. Moreover, I also added the call to the script $sniff\_adapting$, where all the times of the 3 mice are adjusted accordingly to a separate information: during the experiment the A keyboard is pressed and the times corresponding to each mouse have been recorded in that moment. In other words, thanks to this information I was able to adjust all the times in a coherent way, since before that they were coming from separate measurements.
- The function *t_adapting* makes the final adjustment: the 3 times are coherent, but they are still defined on a different number of discretization instants and they overall cover a slightly different interval. After the call to this function, all the 3 mice have activities defined on the same time vector, taken from the greatest common time interval between the 3, and on the same time points (I choose one mouse as reference and interpolated on its time instants for the other 2 mice).
- Some normalization function are available: based on z score, min-max or on the homecage period.
- The function *mice_adapting* takes the data of activities in a mouse and returns a matrix with two columns: the first is the times, the second is the mean of all neurons at that time. In other words this function returns the activity for the all mouse (as discussed in previous meeting, maybe a weighted average could be better, instead of a normal one, tbc in future).
- I also used from now the zone file. With the function *adapt_zone*, for a given activity (always of the observer, either during habituation or test), the correspondent zone file will return a new column where for a specific row (i.e. position (x, y) at time t), it will report the correspondent calcium activity
- I did some plots using the zone file: with *zone_plot* (see Figure), an adapted zone matrix is passed and as result it is plotted the calcium activity colored differently, depending on when the observer was near the

neutral, the stressed or neither. With the plot3 I plotted the calcium activity varying in the space (see discussion for more).

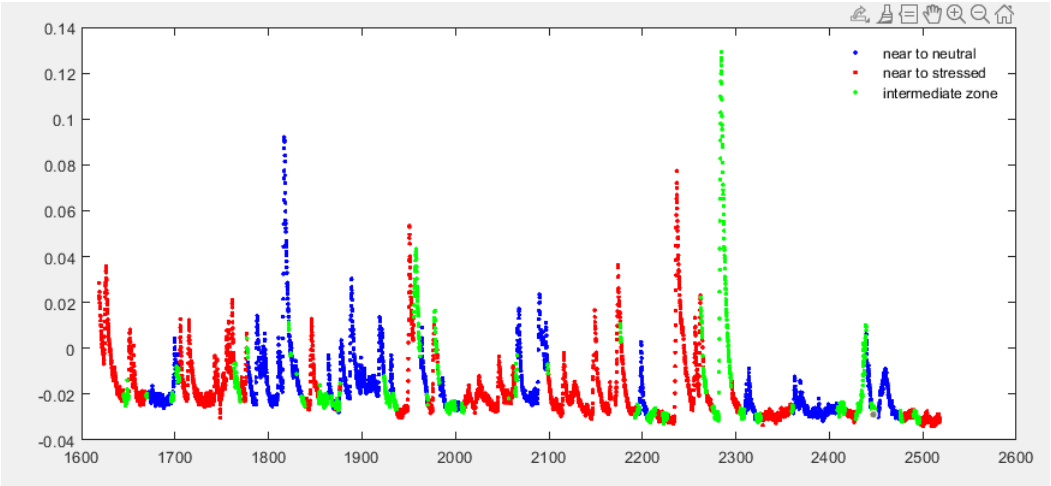

Figure 3: zone plot function's output: different colors for different spatial positions

- Next, I treated the issue of finding a way to establish weather a neuron is active or not, giving as result a binary matrix, 1 if activated or 0 if not at a specific time. A first naive way is using a treshold and considering active everything above it. this is what *activity_detector* does. With *detector_plot*, choosing the neuron you want to see as input, we can have a graphical explanation of this method (see Figure)

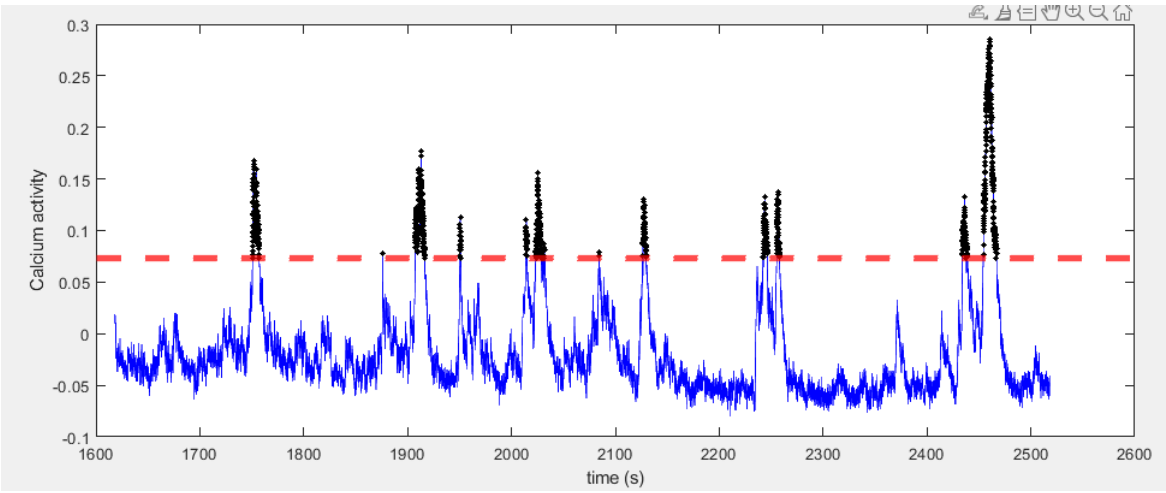

Figure 4: Detector plot function's output: balck dots are neurons considered active. The treshold (in red) is given by $t = \mu + 2\sigma$

- With *mad_tresh* I started to work (not finished yet) on a better tresholding system: I took inspiration from the median absolute deviation (mad) algorithm described at page 67-68 of the Inscopix manual. As we can observe in the figure, this new treshold is varying with the calcium activity, instead of being a straight line

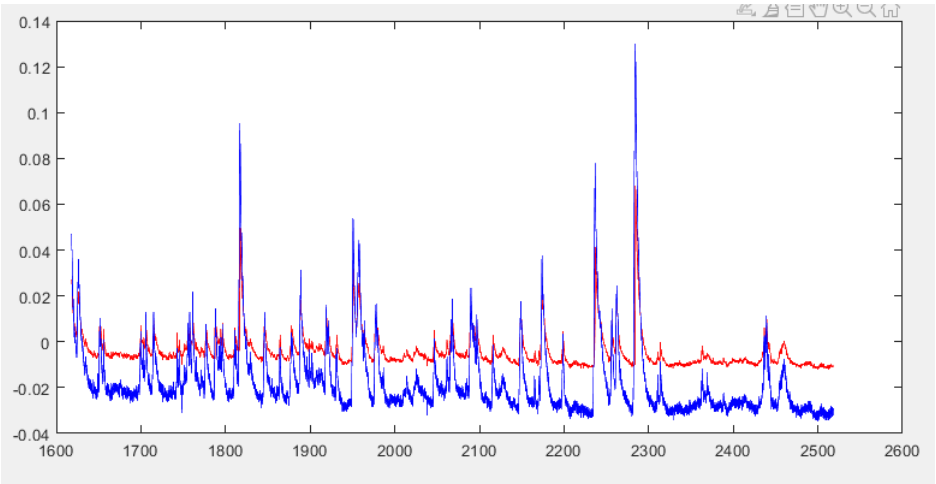

Figure 5: plot mad function's output: the treshold (in red) is given by the curve obtained from mad algorithm

- With the script *histograms* we can observe the histograms of neurons activities in the 3 mice, to observe the activity mean and its distribution around it. it seems that Gamma probability densities (with parameters estimated by Matlab) can well approximate them.
- With the script *mean_and_var* we can see the behaviour of the mean and variance from neuron to neuron
- With the function *plot_correlation* we have a heatmap plot of the Pearson's correlation between the columns of the data matrix passed as input (for example the neurons activities to inspect the correlation between neurons in one mouse)

2.2 Discussion

During the report of the new scripts in Matlab, some comments have been made:

  - We can improve the $3D$ plot of calcium activity vs space. Riccardo suggests to build a mesh covering the spatial boundary and to associate on it the calcium value. Greta suggests the use of a colormap to better capture the variations of calcium. An idea could be to add the time variable and seeing this heatmap in different snapshot over some time instants

- Greta wonders wether the change of directions of mice in the arena could tell us some interesting informations
- The correlation can be applied between two mice, considering their overall activity, to establish how much their activity are linked
- It will for sure be useful to add the sniff data, and relate many of the works illustrated so far with the direct interaction between two mice.

2.3 Next in line

Next work will be focused on:

- Finishing the mad detector algorithm and being able to return a binary matrix
- Improve many of the plots shown
- Work on the 3D plot as discussed
- Including the sniffing informations in the analysis
- As requested by Francesco Papaleo, focus the research on the inter-brain, i.e. how activity and behaviour in one mouse are related to the ones in another, and investigate activity synchronizations.

3 Incontro 3 (27/10/2021)

3.1 What has been done

First of all, I uploaded in the meeting folder in Dropbox, a beamer presentation resuming what has been done up to this point. Mainly, new things consisted in:

- Final time adapting which makes the events coherent with the information on the A key TTL provided in the sniff file: with the *sniff_adapting* script, the times of the 3 mice and of the sniffing informations become coordinated
- I ultimated the MAD treshold algorithm, with the function *mad_detector* it's possible to get the binary matrix corresponding to the signals activations
- With the script *sniff_plot* it is possible to observe the sniffing interactions during the test, marked with dots
- I started a new topic: correlation analysis. Following the Kingsbury article, Pearson and cross correlation have been used as measure of synchronization between signals. The first, however, express the *linear* dependence of two quantities and it could be not particularly fit to our case (in the Kingsbury article, indeed, they show an increase of Pearson correlation, yet an increase leading to less than 0.3 correlation, not very meaningful). The cross correlation, on the contrary, seems the best tool to determine correlation between two signals.
- The two correlation are defined as it follows(in continue and discrete form for the cross one):

$$Corr_P(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

$$Corr_C(X,Y) = X \star Y(t) = X(-t) * Y(t) = \int_{-\infty}^{\infty} X(t-\tau)Y(t)dt$$

$$Corr_C(X,Y) = X \star Y(m) = \sum_{n=0}^{N-m-1} X_{n+m}Y_n$$

In the figure an example of cross correlation behaviour: it is computed as function of the *lags*. The peak of cross correlation corresponds to the delay between the signals, hence a peak near to 0 means that the 2 signals are alligned. Hence, a high (normalized) cross correlation is suggested, beside from an appropriate level between [0,1], from the presence of a possible unique peak corresponding to the allignment of signals.

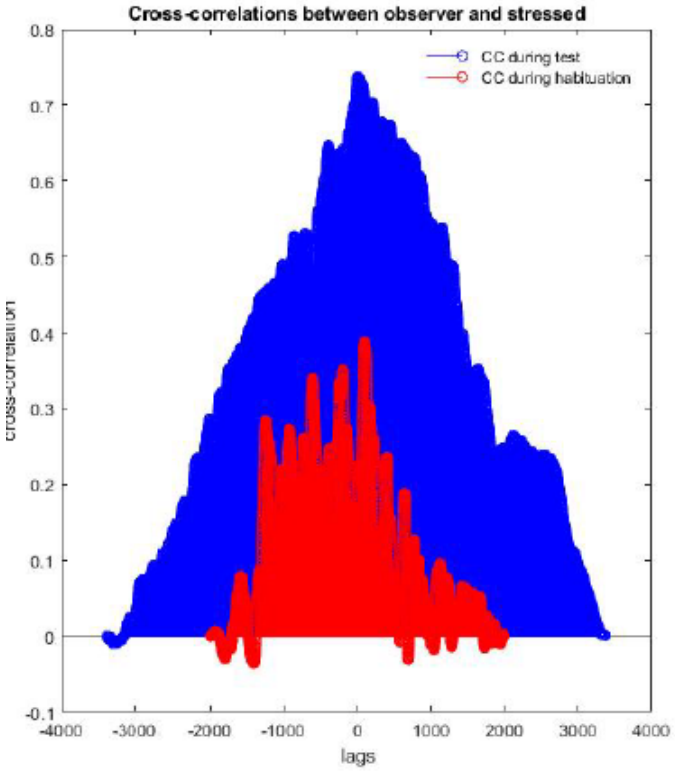

Figure 6: Cross correlation between observer and stressed mice during test (blu) and habituation (red). We can observe a nice behaviour of the correlation for the test, compared to a no correlation behaviour in the habituation. This is reasonable: the correlation during the test is computed over times where observer and stressed were actually close, while during the habituation demonstrator mice are not present, and no correlation should be manifested

- For both the couples observer-stressed and observer-neutral, the test phase shows an increased cross correlation w.r.t the habituation phase. Moreover, if we restrict even more the times of interest to the one when actual sniffing happened, the cross correlation between observer and stressed (but not observer and neutral) is the highest recorded (see Figure)

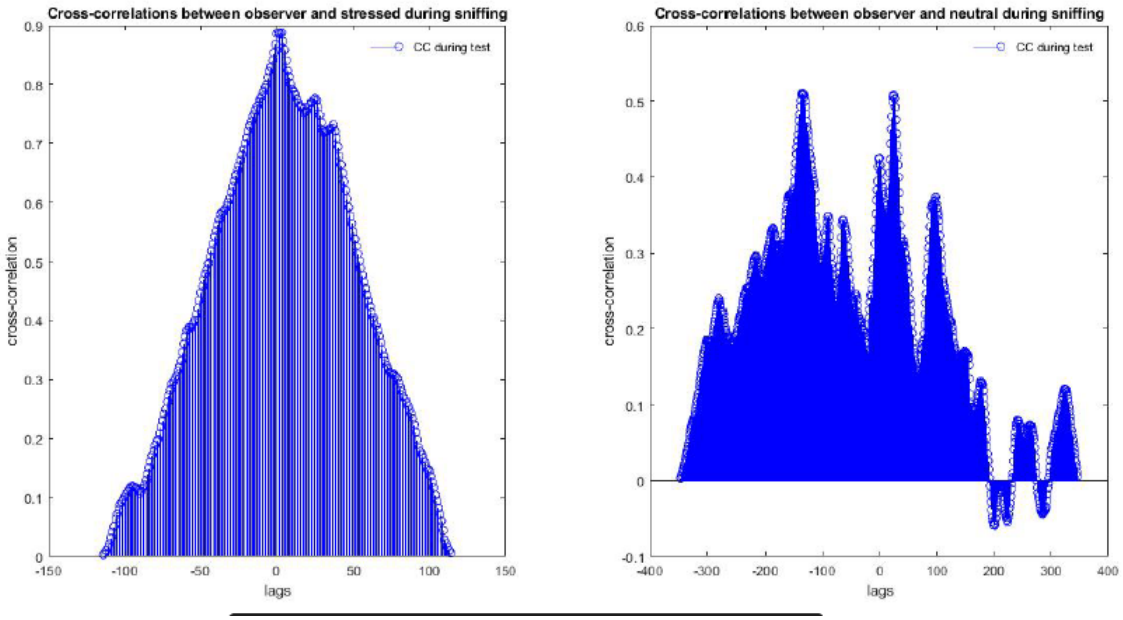

Figure 7: Cross correlation between observer and stressed (left) and observer and neutral (right) during sniffing activities

- The next analysis focused instead on looking for correlations between single neurons activity: for every couple of neurons from one mouse and another, the cross correlation between the 2 signals has been computed and showed in a heatmap matrix. We can identify more clearly the couples showing synchronization coloring only the ones with a correlation above a certain treshold (I selected 0.6). Both for the matrices of the couples observer-stressed and observer-neutral, during the test phase the number of pairs showing correlation is significantly higher than during the habituation (see Figure)

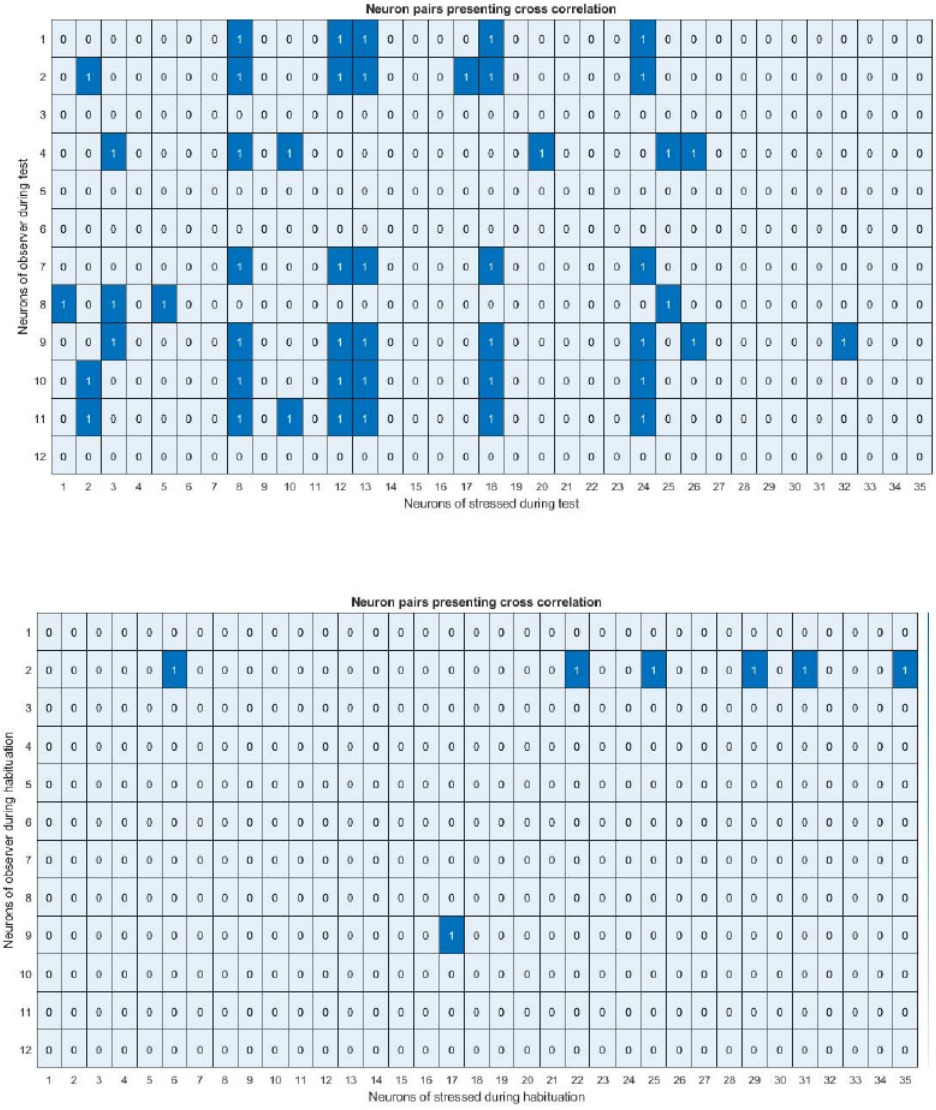

Figure 8: Synchronized pairs during test (top) and habituation (bottom). Similar results obtained both for observer vs stressed and observer vs neutral

- The latter analysis allows also to establish, for every mouse, which neurons are the ones "dedicated" to the correlation. Interestingly, for the observer the same neurons showed correlation with both the stressed and the neutral, even if coming from separate analysis

3.2 Fiberphotometry

Beside the Inscopix procedure to detect single neurons calcium activity, there is a second technique which will be object of work: **fiberphotometry**. A short

guide on what fiberphotometry is can be found at
https://www.mightexbio.com/fiber-photometry/
In short, it is a technique consisting in a light stimultation and light emission in the brain resulting in a collective information on the mouse neural activity. There are not anymore single neurons informations, but only an overall intel, but this technique is faster, less invasive and cheaper. In the following I will start to work on Matlab also for the data coming from it (TBC).

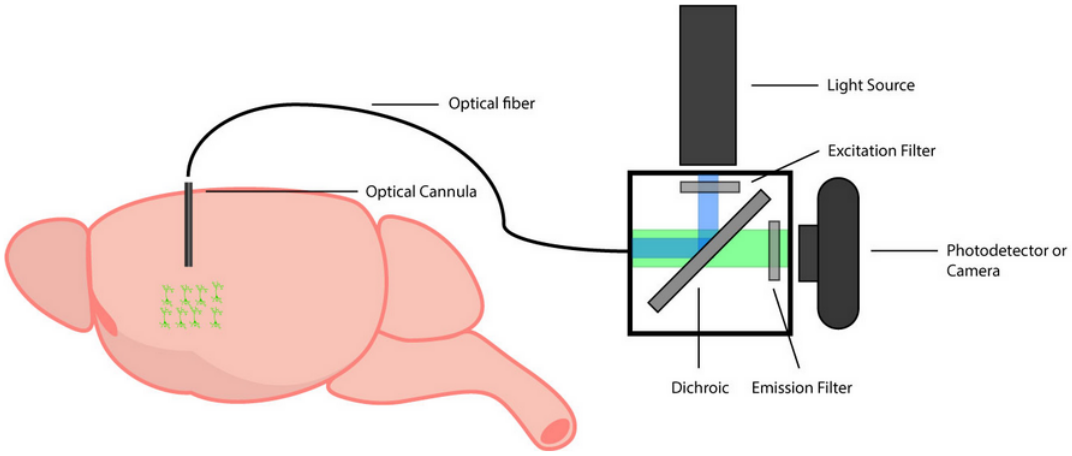

Figure 9: Fiberphotoemtry experiment's setup: see the link for more

3.3 Discussion

The discussion has been focused mostly on the first steps to follow in the differential modeling part, to be added to the data analysis part in the final work. Some discussions on:

- The difficulty in costructing a single cell model for the calcium activity, both for a non sufficient spatial resolution of the Inscopix microscope, both for the difficulty on establish the boundary conditions for a model of that type in relationship with the mouse behaviour during the experiment
- As consequence, it may be wiser to focus instead on the neuron's network, focusing on the calcium activity patterns and electrophysiological links between neurons through axons
- As Riccardo suggests, the network could be threated as a collections of cables pairing single neurons. For every neuron of the network, the Ca activity is given by the experiment, and if we manage to find a link between calcium activity and membrane potential, the first one can be threated as boundary condition for a cable equation problem. A model of this type would then predict the potential drift along the network, and its link with the calcium activity for every single neuron
- Fabrizio suggests a couple of mathematical tools which may be helpful in an analysis of this type:

Quantum graphs (https://en.wikipedia.org/wiki/Quantum_graph)
Differential models on networks (chapter 10 of the online interactive book by Barabasi - Albert may be useful: http://www.networksciencebook.com/chapter/10#network-epidemic)

- Riccardo proposes a 2D model for the occupation probability of the observer in the arena (see the *Sacco_modello2D* file in the meeting folder)

3.4 Next in line

Some IIT researchers asked me for help with Matlab on some codes related to fiberphotomety for thier experiments. Helping them should also help me to get familiar with this new technique.
New data will come and the same correlation analysis will be tested on them, with the strong hope that similar results will be concluded (and start to be closer to an actual scientific result).
Finally, more time will be dedicated on the differential model. For all of these reasons, for the time being I won't have time to go further on the data analysis, altough many things could still be done, they have to be postponed.

4 Incontro 4 (TBC)