pchaUser Manual:

Cleaning noisy trials using BasicAnalysisGui

This document concentrates everything you need to know about the functions' package of basic analysis. It is meant both for people using this package for the first time, and also for programmers that want to change things in it (the expansions are more relevant for those).

The package contains the following functions:

* Analysis1 (m and fig files)
* basicAnalysisGui (m, mlapp and asv files)
* bloodVMask.m
* choose\_polygon.m
* cleanNoisyTrials.m
* createAllcondsMats.m
* createConds.m
* mfilt2.m
* mimg.m
* normalizeToCleanBlank.m
* normalizeToFrameZero.m

this analysis uses codes written by the following people: Inbal Ayzenshtat (2006), Roy Oz (2018), Amit Babayof (2019), Noam Keizer (2020) and Yarden Nativ (2020), and some more contributors who didn't leave any name on their additionals.

This document was written by Yarden Nativ at 22.09.2020, and updated by … at…

**Table of contents:**

Table of Contents

[**1. What the analysis does:** 2](#_heading=h.gjdgxs)

[**2. A few preceding steps to using "basicAnalysisGui":** 3](#_heading=h.1fob9te)

[**3. Steps of analysis** 5](#_heading=h.2et92p0)

[**4. How to choose noisy trials** 7](#_heading=h.3dy6vkm)

[**5. How to run from anew the analysis** 11](#_heading=h.4d34og8)

[**Supplementary 1: the constant variables of different steps and their reason** 12](#_heading=h.17dp8vu)

[**Supplementary 2: statistics behind cleanNoisyTrials** 12](#_heading=h.3rdcrjn)

[**Supplementary 3: some advices for more advanced programming changes:** 13](#_heading=h.26in1rg)

# **1. What the analysis does:**

Basically, the camera files contain a fluorescence map. The files of the VSD recordings (membrane potential) and optical imaging of intrinsic signal (hemodynamic response) are recorded trial by trial, for each one a 100X100 matrix of pixels, each one having one of 256 frames recorded. When we want to analyze this data, we need to do a few basic steps- group all trials by the same stimulus characteristics (called 'condition'), normalize each condition by its baseline so that all measurements all along the recording will be at the same scale, and normalizing all conditions by blank condition (no stimulus) so that we can compare between conditions and detect pixels of blood vessels and pixels outside the chamber limits. Furthermore, some of the trials might be not useful because of several noise sources - for example motion artifact of the animal, crashes/bugs with measurement equipment, some eye movements limitations, dye bleaching, heart beat and breathings artefact and so on. For this reason, before we start using the recorded data, we want to run them through a preprocessing analysis- group them, normalize them and ignore the not useful trials.

For more details please see [Ayzenshtat et al. 2010 figure S12](http://www.jneurosci.org/content/30/33/11232.long).

This basic analysis saves all of the levels of grouping, normalizing and filtering the noises, so that later on, when you make a full analysis, you can always go back to any step and check whether you prefer using a noisy trial that first looked irrelevant or not. For this reason, using the basicAnalysisGui might take you some time, also computational time.

The outputs of the analysis are:

1. conds matrix: the basic condition, in which all trials are grouped together by their stimulus characteristics. The matrix including all data in matlab format before normalizing or cleaning. size: 10000X256XnumOfTrials
2. condsX matrix: the matrix of conds after normalization to the condition's baseline (called 'frame zero'). size: 10000X256XnumOfTrials
3. condsXn matrix: the matrix of condsXn after normalization to blank condition which is without its noisy trials. size: 10000X256XnumOfTrials
4. condsAN: condsXn matrix after deleting noisy trials from it. Noisy trials will appear as mat files under the subfolder 'noisyfiles'.
5. all experiment trials as mat files, one by one, while noisy trials are moved to the subfolder 'noisyfiles'.
6. pix\_to\_remove: blood and chamber pixels saved as masks.
7. optional: meanCleanConds, mean values along trials of condsAN

# **2. A few preceding steps to using "basicAnalysisGui":**

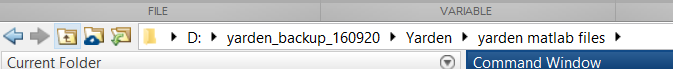
basicAnalysisGui runs on a matlab type files after they get converted from rsd files. So that first preceding step you need to do in order to use basicAnalysisGui is to **make a new folder** in which you want all results to be saved. We recommend on giving the folder a clear name containing the date and session of the file, and write at the folder's name that this is a preprocessed data (for example 'preproccessed\_Merry\_210920a').

After you open this folder, you need to use the GUI of analysis1 to save the full matlab files in this folder. 

**Pay attention!**

1. Analysis1 generates a variable called 'conds' alongside other matrices. Please delete it, because basicAnalysisGui will recreate another one formally.

2. After you've finished using analysis1, please copy to matlab path the path of the new folder with all the matlab files, otherwise basicAnalysisGui will not work.



Next precessing step is to **open a follow-up file** in which you can write and follow all your decisions during basicAnalysisGui, for example you can use this format:

| **Date** | **Session** | **Relevant frames** | **Blank cond** | **highpass filter for blood vessels** | **Noisy trials** | | | | | | | **Made by** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** |

At your convenient, after you finish for each condition choosing the noisy files, the function displays at Matlab's workspace window the names of the files that were moved to 'noisyFiles'. This will allow you to fill the excel file more easily.

You are almost there! Last thing we recommend to do before running basicAnalysisGui is to **check that the analysis is compatible with the kind of signal you are cleaning** (VSD or intrinsic). Please make sure that frame zero variables at the functions normalizeToFrameZero.m and cleanNoisyTrials.m are suitable for your kind of signal:

| Type of signal | Frame zero (baseline) |
| --- | --- |
| VSDI | 25:27 |
| Intrinsic | 5:10 |

**Optional:**

1. In case you want the analysis to save for you output of average among trials of each clean cond- change variable saveMeanResults from 0 to 1.
2. You can see more advanced options at [expansion 3- some advices for more advanced programming changes](#_heading=h.26in1rg).

# **3. Steps of analysis**

1. **Running basicAnalysisGui from command line in workspace:**

There are two ways to run basicAnalysisGui. One of them is to click the mlapp file and run it. The second one, which is more recommended, is to write at the command line the following command:

basicAnalysisGui(xSize,ySize,numOfConds,blankCond,date)

which should look like this:

1. **basicAnalysisGui automatically creates the matrices of conds and condsX:**   
   basicAnalysisGui will first call the function 'createAllCondsMats', which in its turn creates the matrices of conds (all t`rials grouped together by their condition) and condsX (all trials divided to the mean value of the pixel during baseline, aka normalization to frame zero). If any of these two matrices already exists in your folder, createAllCondsMats will skip the part of creating them.

After this step is finished, you will see in your folder the two matrices as mat files.

1. **Cleaning noisy trials from blank:**   
   using the blank condition from condsX, we now clean noisy trials from blank condition. cleanNoisyTrials function's output will be both blankAN which is condsX blank condition without noisy trials, and blank4BloodVMask which is conds blank condition without noisy trials. The variable blankAN will be used for blank normalization (heart beat artifact removal) and blank4BloodVMask for creating masks of chamber limits and blood vessels. In order to clean noisy trials, please read next part, [how to choose noisy trials](#_heading=h.tyjcwt).



1. **Creating masks of chamber border and blood vessels using blank4BloodVMask:** basicAnalysisGui calls the function bloodVMask which makes a matrix mask for blood vessels and chamber border. For chamber border, the function uses constant variables (see [expansion 1](#_heading=h.2s8eyo1)), while for blood vessels the function needs you to choose manually the right filter (cleaning most pixels of the biggest blood vessels). After you found the right filter, please write at your follow-up file what you chose.
2. **basicAnalysisGui automatically creates condsXn using blankAN**: blankAn is one of the outputs of step 3, cleaning noisy trials from blank. Using this variable, basicAnalysisGui calls the function normalizeToCleanBlank, which divides all stimulus trials by the average of the blank condition without noisy trials. This way, we will be able to see the stimulus more clearly, and further on to compare between different conditions.
3. **Cleaning noisy trials from the stimulus conditions:** now we can clean the rest of the conditions from noisy trials. In order to clean noisy trials, please see next part, [how to choose noisy trials](#_heading=h.tyjcwt).
4. **Saving all clean conditions to CondsAN and optionally their average to meanCleanConds**

# **4. How to choose noisy trials**

**Inputs and outputs of cleanNoisyTrials:** the function cleanNoisyTrials uses interactive interface that allows you to choose noisy trials to clear from the data. The function has two outputs- one is a matrix based on the normalized condsXn variable (conds normalized by frame zero and blank condition) that its noisy trials were removed of it. If the function input is the blank condition, the output will clean noisy trials from condsX and not condsXn, so that the clean blank condition will be used to create condsXn. The second output is relevant only for creating the mask of the blood vessels from conds blank condition without its noisy trials, so that this output is empty for any other condition.

Another important thing the function does, is to move noisy trials into the subfolder 'noisyFiles' and display the reasons they were moved/dumped at the workspace. This is a very helpful tool for you to track changes during the cleaning, and write them at your follow-up file.

A useful tip: at any recorded session, it is written in the notebook or in an excel file how motivated was the animal during the session. This could help you evaluate if you're going to have a lot of noisy trials or not, because there's a high correlation between motivation and noisy trials.

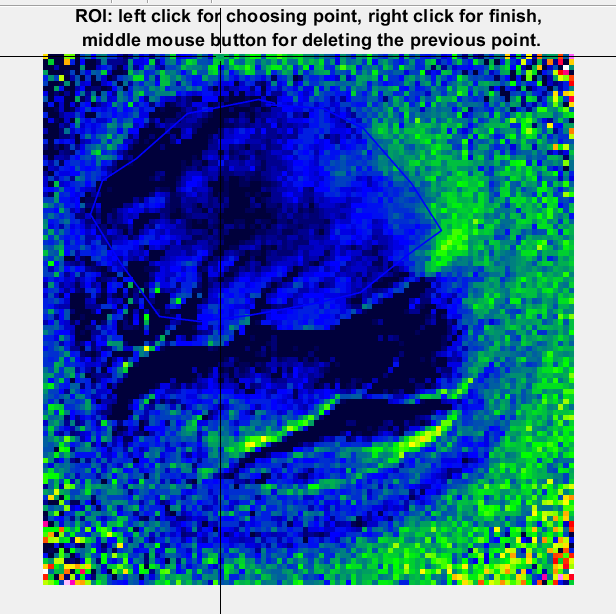
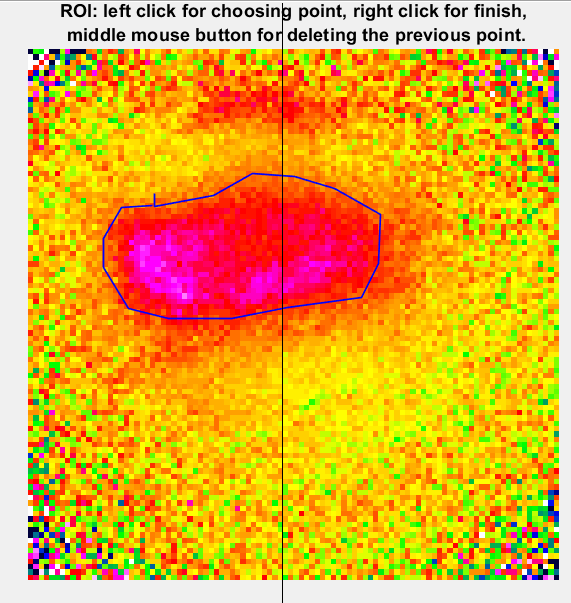
**Levels of choosing noisy trials:**

**1. Choosing frames for ROI:** when you want to choose noisy trials, you need to choose an area in which all statistical analysis will run. Many parts of the camera image might be irrelevant and just very noisy, so you will need to draw the relevant area for cleaning, this area is called 'ROI'- region of interest. In order to choose this area, the function first shows you the frames of the camera between 2 and 100, while the interval between each frame is 10 ms.   
Frequently, the stimulus begins to appear more clearly around frame 35 for at least 150 ms, meaning until frame 50. This is why you have the option at basicAnalysisGui function to choose for some of the conditions to see mean activity between frames 35 and 50, so it can help you choose the relevant frames in cleanNoisyTrials.  
In some cases, the best SNR of the evoked activity appears somewhat later (depends on the effectiveness of the subtract HB model) – choose your frames according to the best SNR.



**2. Choosing ROI:** after you chose the relevant frames, the interface will ask you to choose ROI. When you choose ROI for blank condition, try to choose the biggest area you can with as less activity on it as you can, and try not to include blood vessels- because they make a lot of noise and interrupt the statistics afterwards. It is important to know that although you picked frames for choosing the ROI, the statistical tests will run on all frames (from 2 to 200), but only in the ROI you chose.

Examples for ROIS in blue line (blank conditions on the left):

Some useful tips:

1. For blank condition choose the entire V1 area – including blood vessels. We are interested in the blood vessels noise – as we want to remove it from the further analysis. Moreover – noise in the blank conditions will affect all stimulated conditions (because all stimulated conditions are divided by the blank).

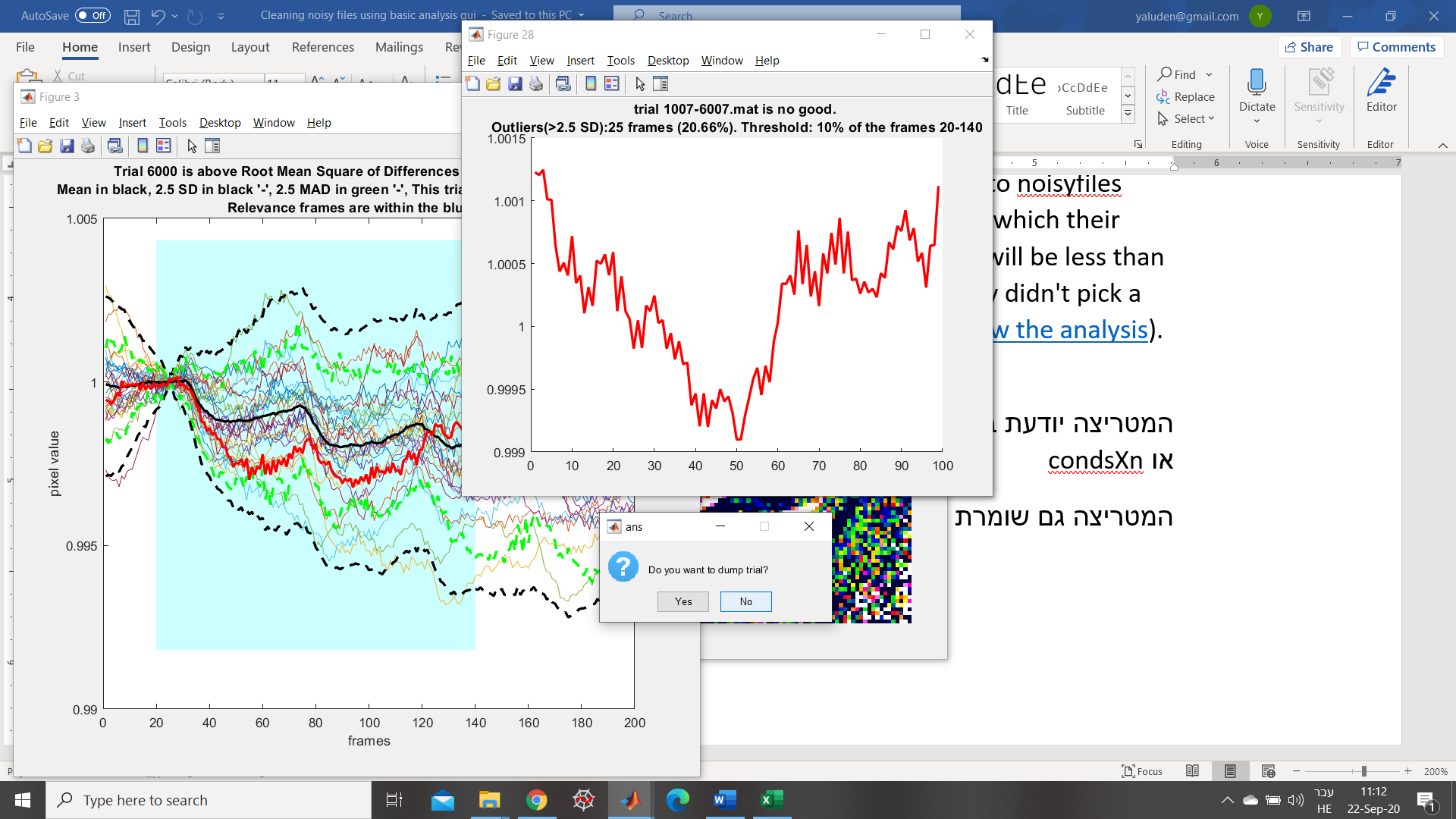
2. For the stimulated conditions choose the ROI with a good SNR of the evoked response. There is a balance between the size of the ROI and the best stimulus response – while larger ROI will reduce independent noise among pixels, it can also dilute the visually evoked response. This is true for small visual stimuli that evoke small regions in V1.

3. Further on in the analysis you might find out that the ROI you chose was not the best. The Basic analysis Gui is built in a way that you can exit all windows and run from anew the analysis for the condition you'd like to fix (see [How to run from a new the analysis](#_heading=h.1t3h5sf)).

4. Based on the purpose of your research, note that the limited ROI also may define trials where the eye position away from the fixation point – as noisy and in fact you might be interested in these trials (for example in research focusing on eye movements). This is the reason we save the noisy trials in a different subfolder and don't delete them, so that later on we can check them again.

**3. cleanNoisyTrials** **automatically dumps outliers:** the function moves to noisyfiles subfolder automatically every trial which has more than 10% frames which their fluorescence value is more than 2.5 SD. Most of the cases the outliers will be less than 3 trials, so that if the function dumps more than that you've probably didn't pick a good ROI and you should start from a new (see [How to run from a new the analysis](#_heading=h.1t3h5sf)).

Those are the two messages you'll get for this automatic process:

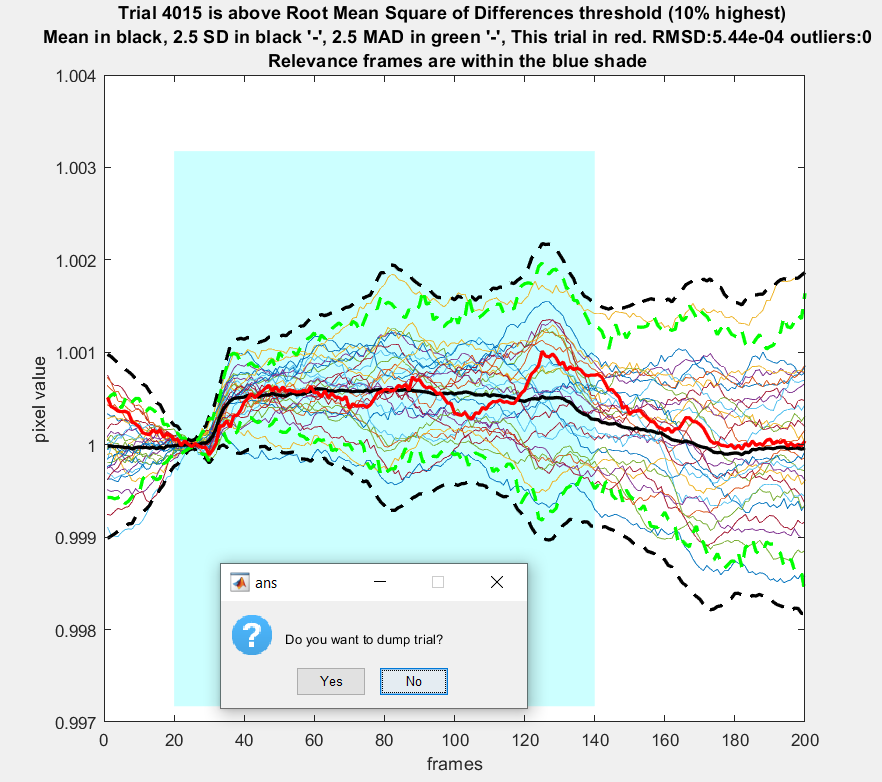
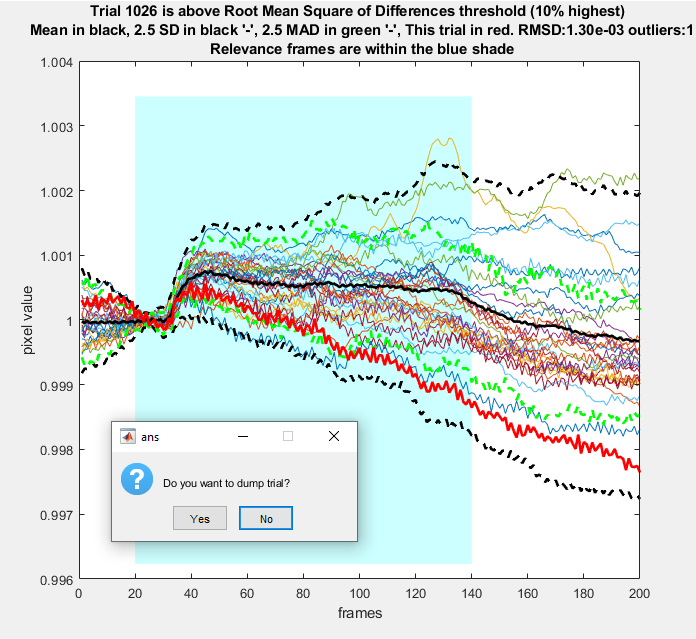


At workspace window: 

**4. Deciding for trials above RMSD threshold:** the function will show you first trials with high RMSD, meaningly trials which their mean distance from the average was averagely very high. This could be connected to a sudden jump in activity later than the average time, to heartbeats or breathing, to big eye movement or to problems with the recording equipment. Useful tip: heartbeats will usually look like a sinusoidal signal above your signal, and this might mean you need to run analysis from a new because you chose ROI with too many blood vessels.

At the window you will see all other trials next to the trial, coloured in red, that you need to decide for. In the window you'll see in black full line the mean values, in black dashed line the limit of 2.5 SD and in grean dashed line the limit of 2.5 MAD (it's like SD, but for median).  
In blue shade, you'll see the relevant frames for your analysis. Sometimes the RMSD could be outside this window of time, and then it's not relevant for dumping.  
Here are two examples for trials, while the first one should not be dumped and the second one should:

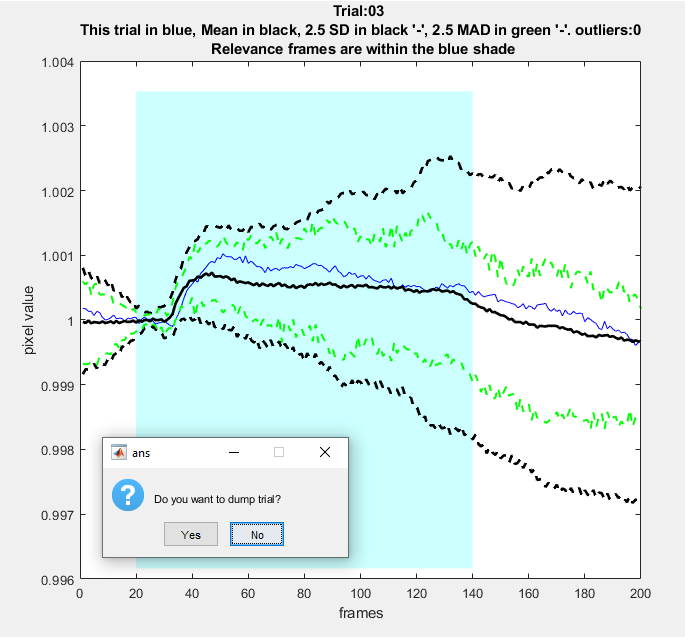
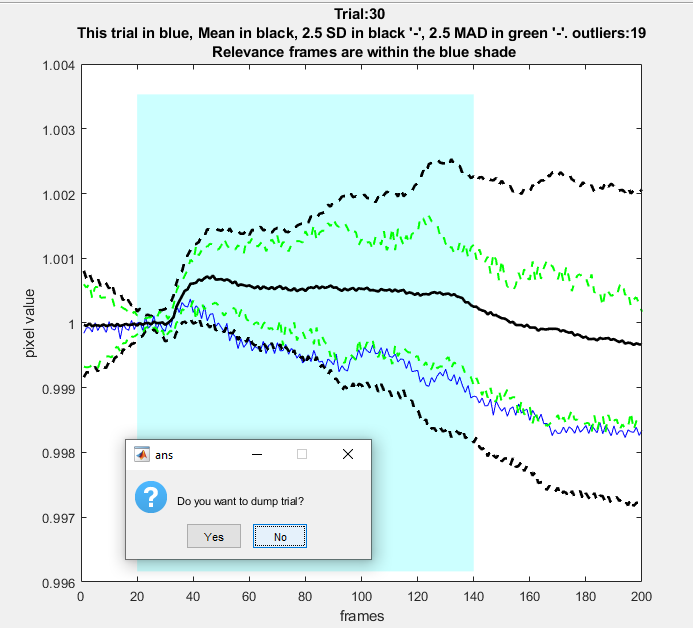
(1) not a noisy trial: (2) noisy trial:

1. **Deciding for all other trials:** now the function will show you the rest of the trials, still showing the mean, the 2.5 SD limits and 2.5 MAD limits. Two things should attract your attention- trials that are outside 2.5 SD or 2.5 MAD at important periods (for example frames 30-50), and trials that their pattern is very different from the mean (for example they go down in activity for a long time while the mean is going up).

Examples:

(1) not a noisy trial: (2) noisy trial:

**Pay attention!** Even if you dumped a trial that later on you'd like to use, you can always bring it back from the subfolder 'noisyfiles' and run from a new analysis so it will be saved in your condsAN file. This is one of the reasons the follow-up file is very important.

Some useful tips for choosing noisy trials:

1. Limit your checkup to the first 500 ms – because later on the animal makes MSs and you may get large variance that you want in and not out.

2. When you write the follow up, please note to yourself noisy trials that needs to be rechecked, for example if you're not sure why the response was late than average, and check their MIMG and eye position. Maybe there is some interesting information.

3. Commonly, averaged response of blank condition should look like a fading sinusoidal signal. Similary, averaged response of stimulus condition shows a transient increase around frame 25, and from then a moderate decrease. If you see a clear different response (like an increasing response in blank condition), check other ROIs, and if you don't find a better one it might suggest that there was a problem in the session or in this specific condition.

**Pay attention!** In some cases, after basic analysis, there will be more advanced levels of cleaning data using more sophisticated analysing tools such as PCA to find trials with highest common variance. If you need to clean again the trials, or do it in a more advanced way, it means nothing about the quality of your cleaning so far.

# **5. How to run from anew the analysis**

Sometimes, you'll find out that you want to run from anew the analysis- because you decided to change your mind about ROI, or about noisy trials in one of the conditions, or because you have to stop in the middle of the cleaning. Due to the fact the analysis saves all things you were doing in a mat files, and loads the mat files from a new every time, the analysis might take some computational effort and you might wait. On the other hand, every time you run from a new the analysis, it will skip steps of recreating your mat files from a new- meaningly the conds and condsX mat files.

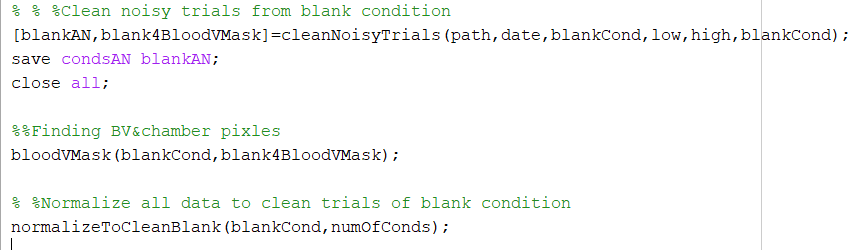
If you want to change the cleaning of the blank condition, you have to run from a new the analysis as you ran it for the first time, because than it might change all of your decisions regarding other conditions (because you normalized by the clean blank).

If you want to change only stimulus conditions, you need to do three things:

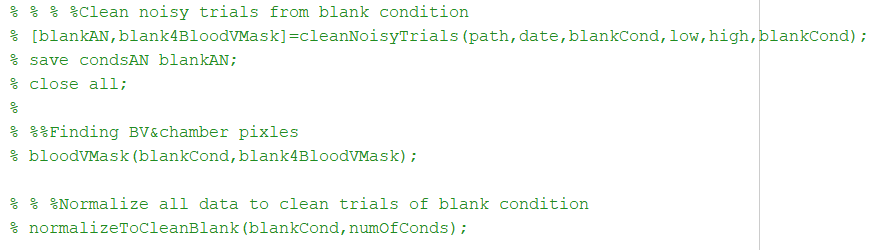
1. Move the files from subfolder 'noisyfiles' of the relevant condition back to the folder with all trials. The camera files are saved in a way so that the date appears first, and number of stimulus and trial last. For example a file named '1007\_2013' is trial number 13 for stimulus no. 2 from the date 10.07.

**Pay attention!** This is a very important step, because sometimes the analysis might use different range of samples to calculate the statistical values, and will dump automatically trials that are not noisy at all.

1. Enter m file of basicAnalysisGui and put under a comment the following lines:



Under comment:



1. Run the analysis from command line using only the conditions you want to clean. For example, if you want to clean only conditions 3 and 4, you should write in the command line:



# **Supplementary 1: the constant variables of different steps and their reason**

**Chamber's threshold of illumination:** In the function bloodVMask we use a minimum threshold of illumination, that above it we define as out of chamber area. The constant of threshold is 15% out of the maximum activation of the camera- 16384. The constant is connected to the Fluorescence ratio for VSDI (In most cases the brain has higher fluorescence ratio and higher reflection ratio than the chamber/dura).

**2D High-pass filter for detecting blood vessels:** In the function bloodVMask we use the function mfilt with its constant variable for 2D-high-pass filter to see blood vessels. The filter size (XX, 'hm') is the spatial width of the filter. Blood vessels are considered to have high spatial frequency (the typical width of a blood vessels is only few pixels, within the imaged frame). The high pass filter removes the low spatial frequencies within the chamber and keeps only the high pass features which are mostly the blood vessels and the artificial dure etc. When we use high-pass filter we can see the blood vessels, and then pick how much of them we want to mask. As you can see in the images there are a lot of blood vessels in the area the camera recorded, but we don't want to remove too much of the pixels- so we focus mainly on the big blood vessels and remove them.

**Frame zero for VSD signal:** in the functions cleanNoisyTrials and normalizeToFrameZero we use frames 25-27 as baseline frames for normalization (frame zero). We calculate the mean activation in those frames, and then divide each frame by this value for the rest of the trial. This way all trials have the same scale, and we can average between them later when picking noisy trial. Frames 25-27 are the last 3 frames the camera takes before there's a stimulus for the animal on the screen, but it's already after the animal had her pre-cue towards the stimulus. So we can count on the fact the animal is now in a 'ready' position towards a stimulus, but we can use frames which don't have yet any stimulus appearing on them.

# **Supplementary 2: statistics behind cleanNoisyTrials**

(Based on explanation by Noam Keizer, 2020, using his corrections to statistical problems in the code)

Clean noisy trials algorithm is based on three levels:

1. automatically dumping trials with more than 10% outliers (more than 10% of the frames cross 2.5 SD). Outliers are calculated only for relevant frames (see [supplementary 3](#_heading=h.26in1rg) for more details)
2. Deciding for trials with high RMSD (10% of the RMSD values), while RMSD means root mean squares of differences.
3. Deciding for the rest of the trials.

The 2.5 SD lines in the graph is calculated based on the mean value of all trials, and is aligned with the mean value. The 2.5 MAD lines, representing the equivalent to SD for median value, so that they are aligned with the median value which does not appear in the graphs. MAD is given by . The MAD is more suitable to examine extreme trials because it is not affected as much by the mean of other trials as the SD, but mainly by the median.

# **Supplementary 3: some advices for more advanced programming changes:**

dear programmer fellow, we ask you two things in case you want to change some more advanced features of the code:

1. Please write your name in the documentation of basicAnalysisGui and other called functions you call from there and write what was the update.
2. If you need, please fix written things in this document, so that it will be suited with your changes- even if you think you are the only one who's going to use your new script codes.

**More advanced options you can change in the code:**

1. **Using more the one blank condition in normalizeToCleanBlank and bloodVMask:** you can use Noam's additional to both codes in order to use more than one blank condition for the analysis (ICMS exp.).
2. **Changing Relevant Frames in cleanNoisyTrials**: As default, the relevant frames chosen for analysis are 20:140. If you want to change it, please check that you also change in all of shade and plot commands the limits, because otherwise it will show too much or too less data, which will not be fitted to the statistical calculations of the relevant frames.
3. **Automatically dumped outliers in cleanNoisyTrials:** you can change easily the SD threshold for outliers using the variable meanplus2sd, but if you want the function to ask you first before dumping those outliers you'll need a more advanced programming. You can try and use some of the scripts written under the part of derivative, from which you can add in some features that help making this part manual instead of automatic.
4. **Displaying names of files dumped:** the conversion of the camera files to matlab using analysis 1 saves the files using 0 as first trial, and not 1. This is why the display messages now contains the last 4 characters of each file name, so that it will be suited with the camera names and not with the number of trial that might be confusing (1 or 0 for first trial). You can change the display message, but please add a note in this document file or in the code itself, so that it won't be confusing for another user.

**Beware of re-normalizations!**

Many people contributed to this analysis, and this is why it is a very rich one, but on the other hand some of the normalizations were recomputed again and again using different functions inside the code. Before you are doing a normalization, please check twice that you are not re-normalizing the data (otherwise the normalization cancels), and that you are doing it in the right steps as described in [steps of analysis](#_heading=h.2et92p0).

**Efficiency of the code:**

Due to the fact the analysis saves and loads the matrices of conds, condsX, condsXn, condsAN and optionally meanCleanConds, the code might take a lot of time to run at first running. For this reason, you might want to make the efficiency better for saving or unsaving some of those files. Please note that conds is crucial for bloodVmask and for this reason is used in cleanNoisyTrials while used for blank condition. Also condsX is crucial for creating condsXn which will be used for all stimulus conditions in cleanNoisyTrials.

After using the tic and toc functions of matlab, it was found out that saving to mat files without using 'append' is faster. This is the reason why conds, condsX and condsXn are saved this way. For condsAN, to allow the user to stop in the middle of analysis, or run from a new a condition, the 'append' command for saving was used.

Using profile function in matlab, didn't raise any other bold efficiency problems. Nevertheless, you are more than welcome to try and make this code even more efficient, but don't forget to write your updates both in the documentations of the code and in this file.