Beginner's guide to the complete Liverpool SPN catalogue

Alexis D.J Makin and John Tyson-Carr

Contents

1) Finding and downloading data
2) Visualizing and analysing data in Matlab
3) Analysis in eLife paper
4) The SPN App
5) BIDS format

1) Finding and downloading data

When you first click on **The complete Liverpool SPN catalogue** (https://osf.io/2sncj/) you will see a page with components listed on the left (Figure 1). This beginner's guide is in the **SPN catalogue users guide and summary analysis** folder (https://osf.io/gjpr7/).

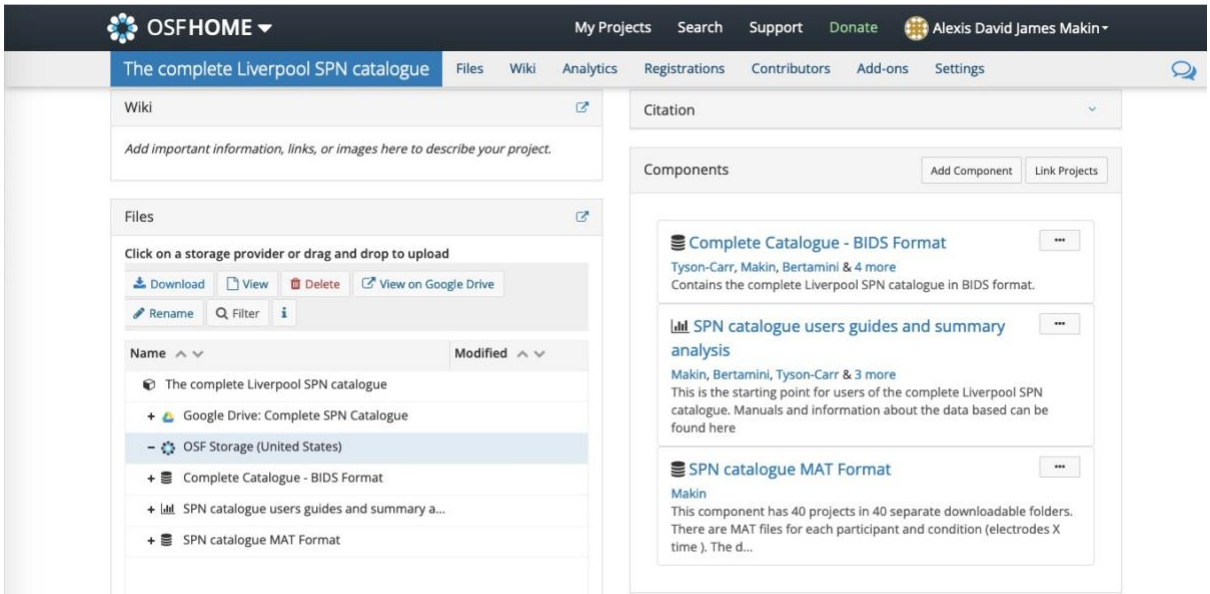


Figure 1 – Front page of the complete Liverpool SPN catalogue

Google Drive – Complete SPN catalogue has over 900 GB of data. There are 40 projects. Each project has two folders (e.g. Project 1 and Project 1 allData, Project 2

and Project 2 allData…). The raw data is in allData. The allData folders have subfolders for each experiment, and within these, subfolders for each participant.

Figure 2 shows the folder structure and basic organization of the SPN catalogue

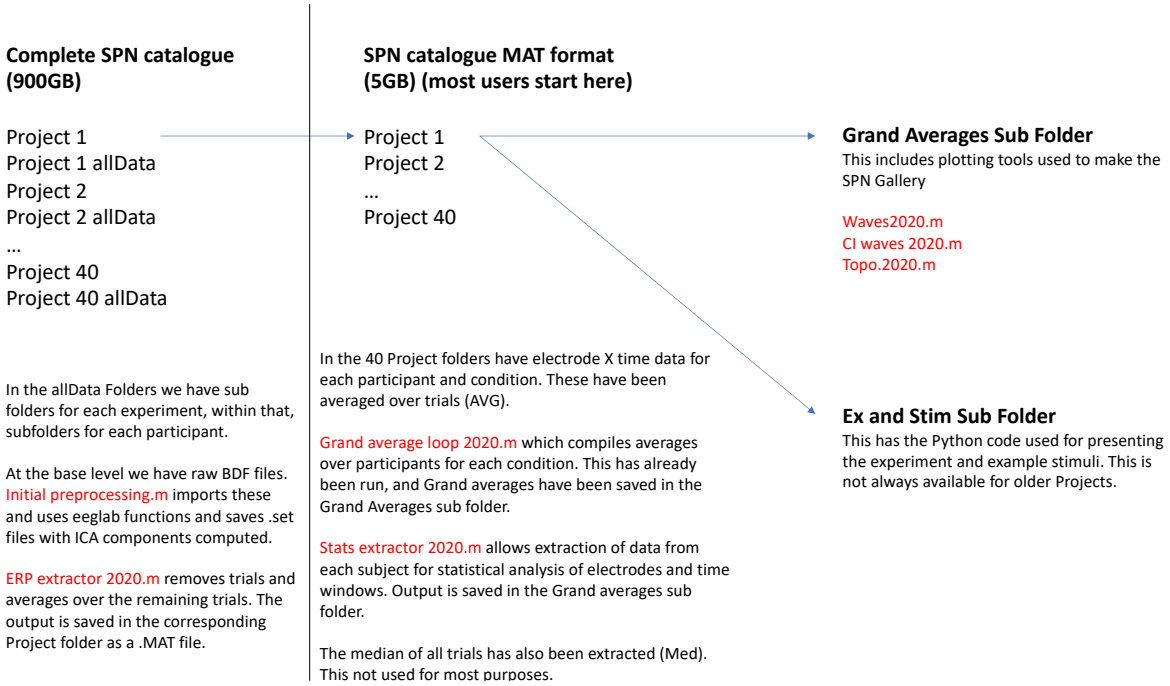


Figure 2 – basic organization of the SPN catalogue

SPN catalogue Mat format

The 40 Project folders are compressed in and available in SPN catalogue MAT format https://osf.io/s4n5b/.

Here there are 40 downloadable projects each has a separate ZIP file. If you download all 40 of these, it will take up 5GB of hard drive space. For many purposes, users may want to download just one Project folder.

2) Visualizing and analyzing data in Matlab

To visualize and analyse this data as described in this section, you will need Matlab installed. You will also need to download eeglab and save Matlab search path to eeglab. This guidebook presumes some familiarity with Matlab and eeglab.

In each project folder, there are MAT files for each participant and condition (electrodes X timepoints). The data has been pre-processed (e.g. referenced to scalp average, low pass filtered, cleaned with ICA and trial removal, and averaged over trials).

The Grand Averages folders have data averaged over participants for visualization of Grand average waves.

There are 40 project folders. Within some folders, there are subfolders for each experiment. For example, in Project 8, there are sub-folders for Experiments 1-5. That is because project 8 had 5 experiments, with 22 participants in each. This project is published in the included PDF (Makin et al., 2016).

Within each subfolder, you can see EEG data for each participant and condition as .mat files. When you click on one of these, it opens an electrodes X time points array called **condAVG.mat**. There were always 64 electrodes, however, the length of the epochs varies across projects. You can open the **timeVector.mat** file to see time points corresponding to each column.

The data has already been pre-processed in Matlab using eeglab 13.4.4b. Typically this involves re-referencing to scalp average, low pass filtering at 25Hz, down-sampling to 128 HZ, and cleaning with ICA to remove blink and large eye movement artifacts. Trials where amplitude exceeds 100 microvolts at any of the 64 scalp electrodes were excluded. Finally, the data was collapsed across trials. The data here is the result of this pre-processing.

The most important thing about this data is that we have already averaged over the trial dimension, leaving electrode X time for each participant and condition.

Next, we consider 6 scripts. **Waves 2020.m, Topo 2020.m, Stats extractor 2020.m, Thin stats extractor 2020.m, Peak extreme stats extractor 2020.m** and **Thin peak extreme stats extractor 2020.m**. Reference to line numbers below are approximate.

The '2020' refers to the fact these scripts were written in 2020. However, they are basically the same as those we have been using since 2014. Scrutiny of these scripts is essential for establishing the validity of our work.

1) Plotting an example SPN with Waves 2020.m

Download project 8 and save it in a location on your computer. Navigate to subfolder EX5 AS4F.

Open Matlab and set Grand Averages folder as your current directory.

In EX5 AS4F, there is a Grand Averages folder. These grand averages can be used to plot ERPs (using the Waves 2020.m code) or topoplots (using the Topo 2020.m code). The grand average data was generated by Grand average loop 2020.m. You will not need to run this again.

Open Waves 2020.m. On line 5, you can see this line.

```
% DIRECTORY (THIS WILL NEED TO BE CHANGED TO A DIRECTORY ON YOUR COMPUTER)
```

```
cd '/Users/alexismakin/Desktop/Alexis Work/Symmetry EEG/SPN effect size and power/Project Folders/Project 8/EX5 AS4F/Grand Averages'
```

This needs to change to reflect the directory where you are working on your computer (or commented out if you are already in that directory).

Then select and evaluate Waves 2020.m. You should get two figures, which look like Figure 1:

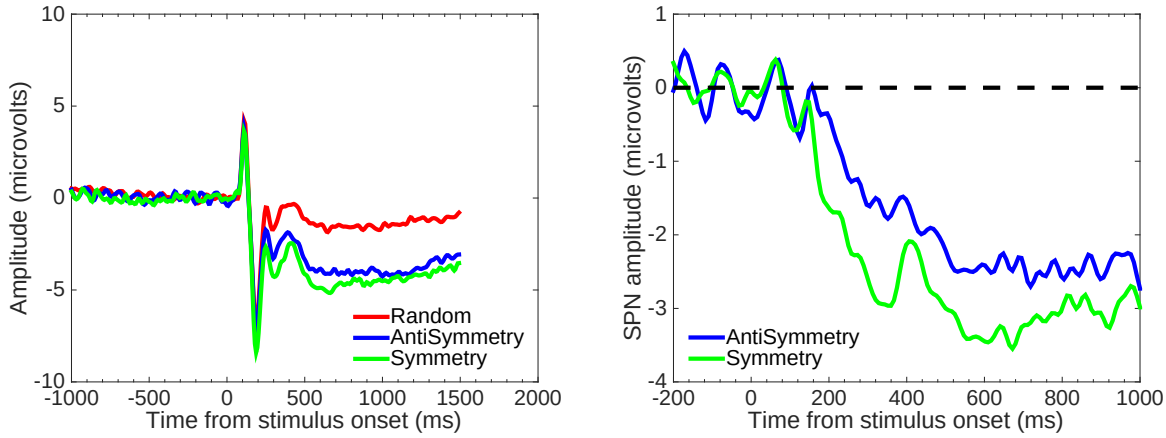


Figure 1: The left shows ERP waves, and the right shows SPN as difference from the random wave at electrodes PO7/8

These were plotted from electrodes 25 and 62 (PO7 and PO8) in the BioSemi montage, as seen in Figure 2. Channels.xyz is a text file which lists channel names, numbers and 3D locations (the coordinate coding system differs from that used in alternatives electrode location file, such as the channelLocs.mat, but the topoplot software knows how to read either).

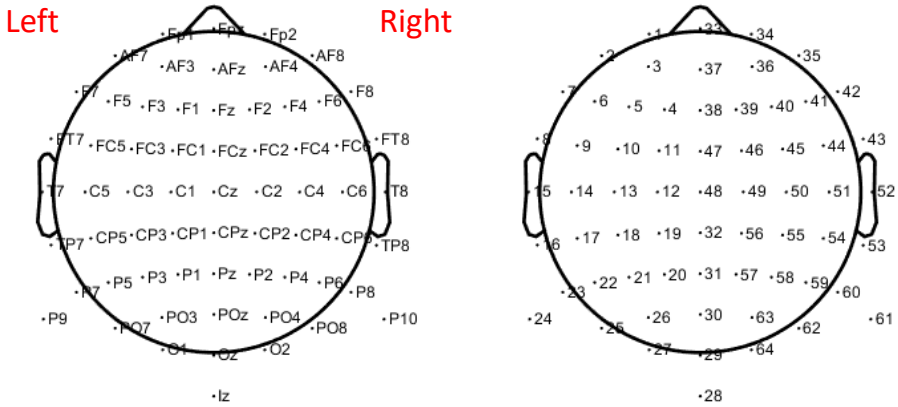


Figure 2: Identity and location of the electrodes.

One of the first things to try is plotting alternative electrodes. This is done on line 12 of Waves 2020.m.

```
%ELECTRODES
electrodes = [25 62];
```

If you enter different numbers in the brackets, you will get different channels. For instance, you might want to try the alternative cluster [PO7 O1 O2 PO8] which is also common in SPN research (referred to as cluster 1).

```
%ELECTRODES
electrodes = [25 27 62 64];
```

Now the waves look slightly different, as you can see on close examination of Figure 3

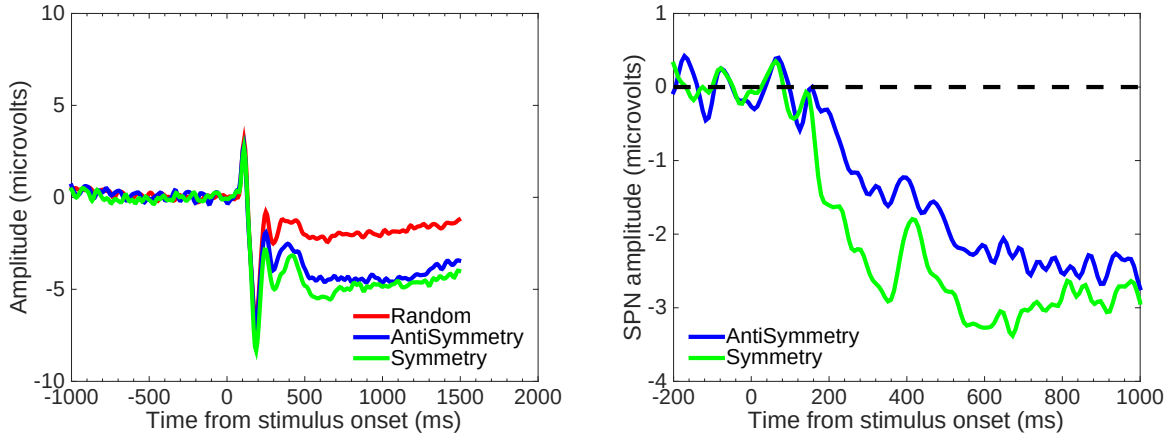


Figure 3. Grand Average ERPs and difference waves at electrode cluster [PO7, O1, O2 and PO8]

Some EEG researchers like to smooth the waves using a moving average filter to smooth EEG waves. Lines 15-17 allow you to turn this on 'on' or 'off' (as in this example) and change the size of the moving average window. You can make the waves smoother by turning this on, but it is a cosmetic thing to do. You can see this in Figure 4

```
%DO YOU WANT TO SMOOTH THE ERP WAVES?
smoothfactor = 10;
smoothOn = 'off';
```

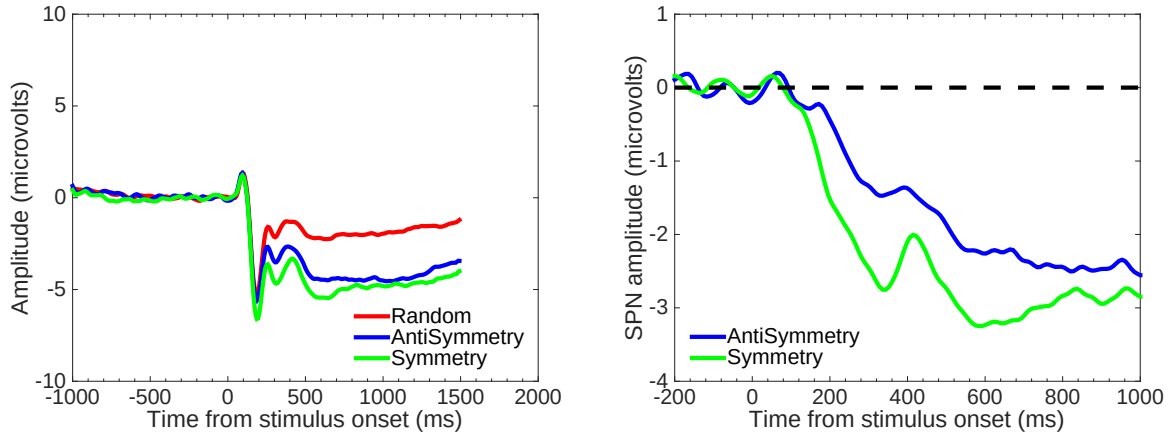


Figure 4. Same waves with 10 point moving average window.

Graphical parameters are coded in lines 17-20. You can play with altering font size and line width here.

Now let's look in more detail at the code the creates the figure. Some of this is rather self-explanatory once you know where to look.

```
%Figure 1
StyleArray = {'LineStyle'};
StyleOrder = {'-','-','-'}';
ColorArray = {'Color'};
ColorOrder = {[1 0 0],[0 0 1],[0 1 0]}';
figure('color',[1,1,1])
P                          =                          plot(timeVector,
[Random,AntiSymmetry,Symmetry],'LineWidth',LineWidth);
axis([-1000 2000 -10 10]);
set(gca,'YTick',-10:5:10);
set(gca,'XTick',-1000:500:2000,'XMinorTick','on');
set(P,StyleArray,StyleOrder,ColorArray,ColorOrder)
legend({'Random','AntiSymmetry','Symmetry'},'FontSize',FontSize,'Loca
tion','southeast');
xlabel('Time from stimulus onset (ms)');
ylabel('Amplitude (microvolts)');
grid('on')
legend('boxoff')
```

Try changing the xlabel from '*Time from stimulus onset (ms)'* to something else for a start.

There are 3 lines in this plot, Random (Red), Anti-symmetry (Blue) and Symmetry (Green) ERPs. You can change the colour of the three by altering the RGB values in the ColorOrder parameter on line 51. Why not try making the Random wave black by setting RGB to [0 0 0] or changing line style to dot-dash? The code now looks like this.

```
%Figure 1
StyleArray = {'LineStyle'};
StyleOrder = {'--','--','--'}';
ColorArray = {'Color'};
ColorOrder = {[0 0 0],[0 0 1],[0 1 0]}';
figure('color',[1,1,1])
P                                =                                plot(timeVector,
[Random,AntiSymmetry,Symmetry],'LineWidth',LineWidth);
axis([-1000 2000 -10 10]);
set(gca,'YTick',-10:5:10);
set(gca,'XTick',-1000:500:2000,'XMinorTick','on');
set(P,StyleArray,StyleOrder,ColorArray,ColorOrder)
legend({'Random','AntiSymmetry','Symmetry'},'FontSize',FontSize,'Loca
tion','southeast');
xlabel('Time from stimulus onset (ms)');
ylabel('Amplitude (microvolts)');
grid('on')
legend('boxoff')
```

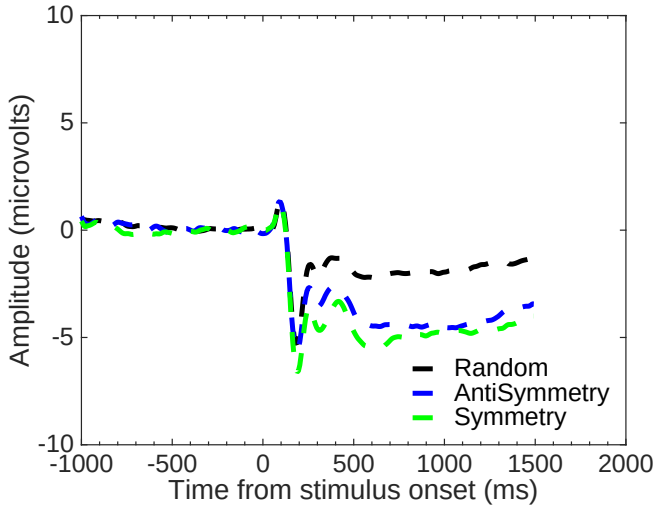


Figure 5. Same waves with 10 point dashed lines and black random wave.

Lines 41-43 are important. These create the difference waves used to plot SPN.

DiffAntiSymmetry = AntiSymmetric-Random;
DiffSymmetry = Symmetry-Random;

Note that we could turn these the other way up, and talk about relative positive deflection in the random condition.

```
DiffAntiSymmetry = AntiSymmetry-Random;
DiffSymmetry = Symmetry-Random;
```

```
DiffAntiSymmetry = Random- AntiSymmetry;
DiffSymmetry = Random-Symmetry;
```

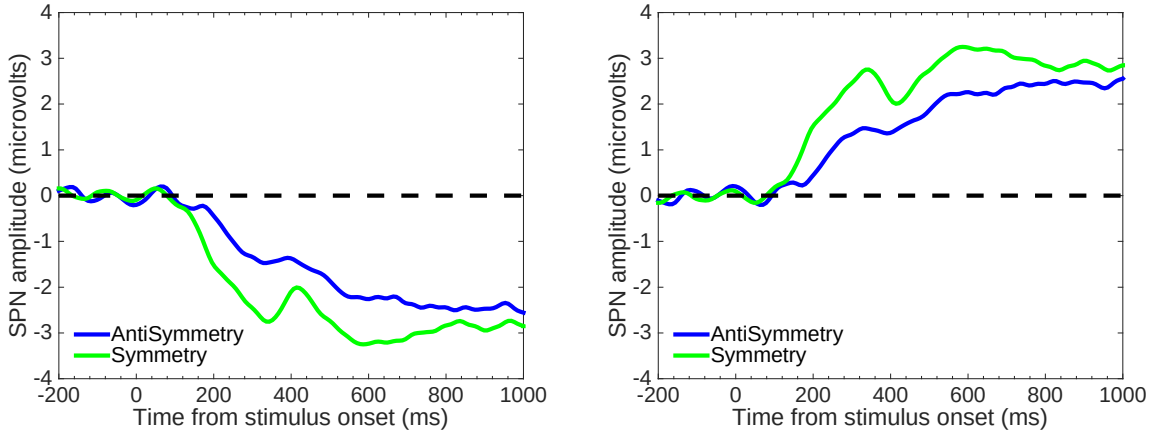


Figure 6: Difference waves produced from code above. Only the left convention is ever used. However, it helps understand the script to look at both.

The dimensions of the x and y axis, and the position of the tick marks, can be set in lines 56-58. In the next example, the x axes were set at -200 to 1000.

2) Topoplots with Topo 2020.m

Next, we consider how to plot an SPN topography, averaged over a particular time window.

Often in our research we plot topographic difference maps, showing the different from random condition at each of the 64 electrodes.

Open Topo 2020.m and change line 5 so it points the directory on your computer (or just comment this line if Matlab is already at the right directory).

This code also reads files from the eeglab library. You will need to configure Matlab so eeglab is on the Matlab search path. If you already use eeglab you may already have this setting saved. If not download eeglab from https://sccn.ucsd.edu/eeglab/download.php and set path to this folder.

On lines 13-15, you can see the time windows used for this analysis. The topoplots use data averaged across this window. If you wished to plot topographies from the 300 to 1000 ms, then these are the right parameters

```
%TIME WINDOWS
low = 300;
high = 1000;
```

Below this, you can see three lines under the header ‘GRAPHICAL PROPERTIES’.

```
%GRAPHICAL PROPERTIES
FontSize = 20;
zscale =3
contourRes = 3;
```

FontSize parameter is important when adding labels.

zscale determines the limits of the colour bar axis in microvolts. If you set this very high, then we zoom out and all topoplots look midgreen.

contourRes determines the number of contour lines on the plot. 3 is quite low but might be appropriate for small topographies.

If you run the Topo 2020.m, you get two topoplot figures, one for anti-symmetry and one for symmetry (Figure 7). They are saved as PNG image files in the Grand Averages folder, because these copy-pasting the Matlab figs into PowerPoint or a graphic software does not give a high-resolution image. It is much better to use the PNGs.

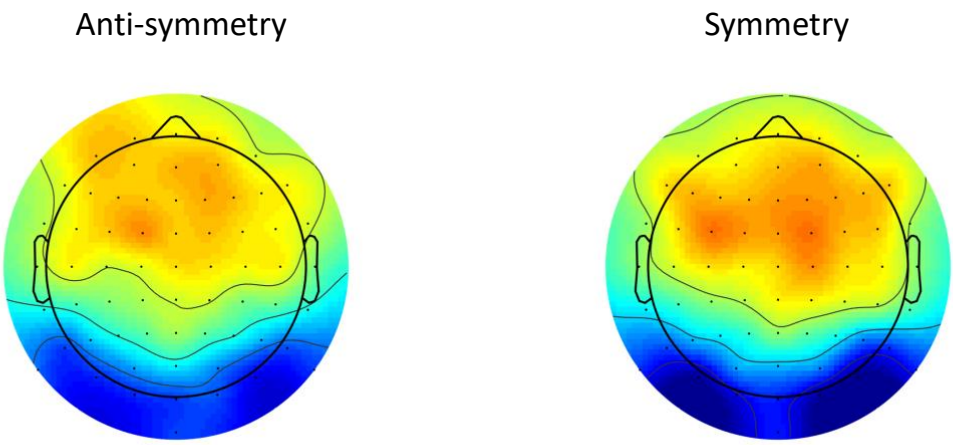


Figure 7: Anti-symmetry and symmetry topoplots made in Waves 2020.m

These topoplots in Figure 7 are quite sparse. They can be supplemented by uncommenting colorbar and title lines 42 and 44 and generate topographies like those in Figure 8.

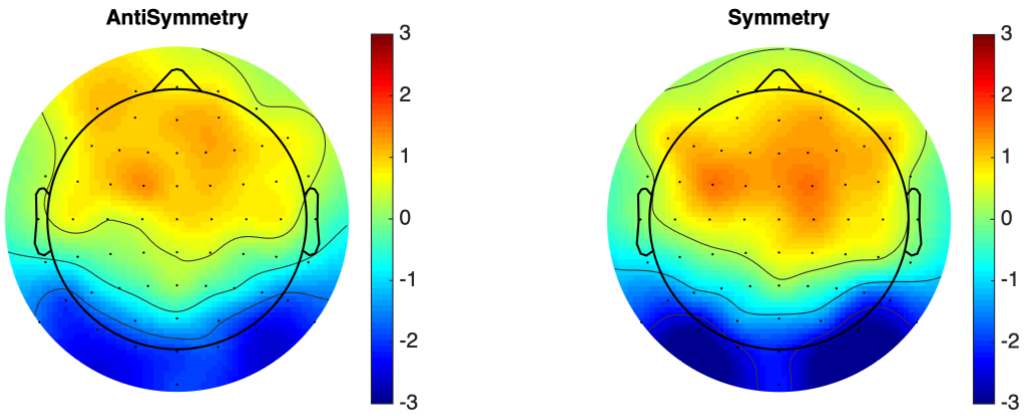


Figure 8: Anti-symmetry and symmetry topoplots made in Waves 2020.m, this time with labels and colorbar

You can see the consequences of changing the zscale to 10 in Figure 9. Now all amplitude fluctuation is in the mid green range.

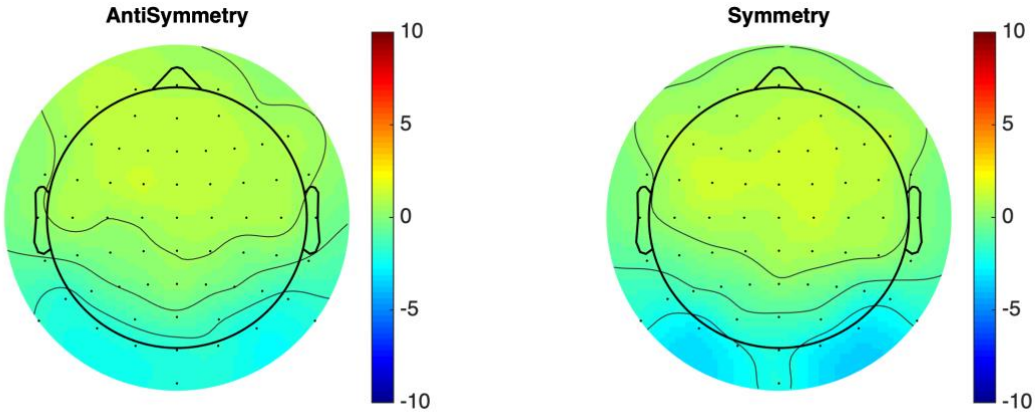


Figure 9: Anti-symmetry and symmetry topoplots made in Waves 2020.m, this time with zscale set at 10.

Figure 8 and 9 show topographic difference maps. On can also edit Topo 2020.m to plot the ERPs rather than differences, and to invert the polarity of the difference map (Figure 10). This can be done by changing line 38 from `DiffSymmetry =Symmetry-Random;` to `DiffSymmetry =Random-Symmetry;`

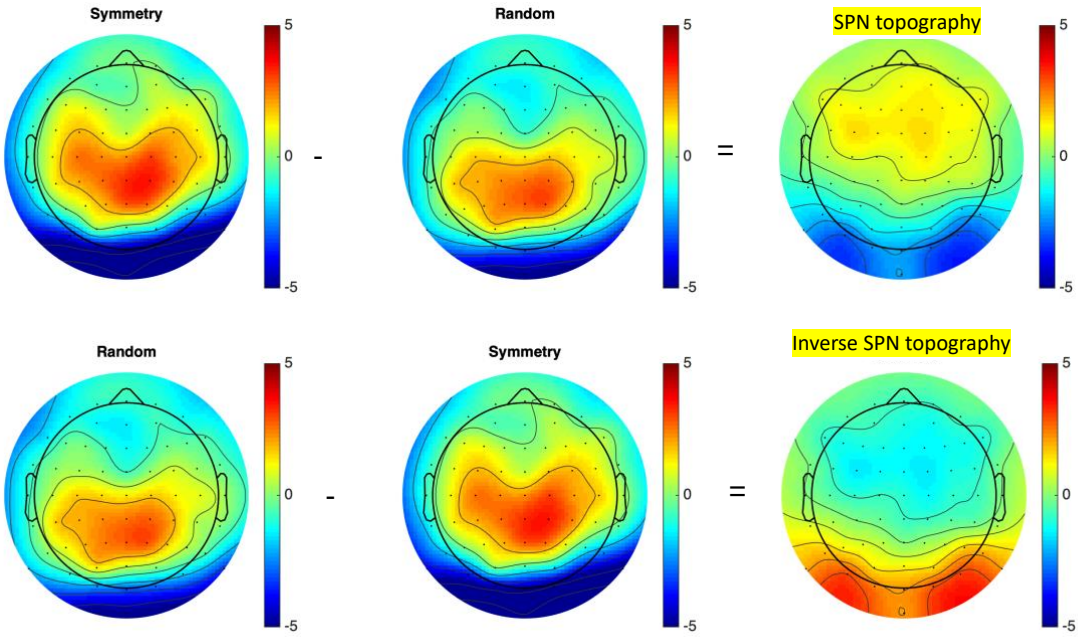


Figure 10. Topography in symmetry and random conditions (300-1000 ms) and the difference map. Normally we plot symmetry – random, but one could also plot random – symmetry to see the same difference map with inverted polarity.

The numbers used in these topoplots are shown in Table 1 on the next page. You can get these from the Matlab workspace after running the topoplots. This can be useful for running some basic forms of source localization on the grand average data.

Electrode	Symmetry	Random	Diff
Fp1	-0.2354	-0.8422	0.6068
AF7	-0.6681	-1.1861	0.5180
AF3	-0.6145	-1.3866	0.7721
F1	0.3992	-0.7460	1.1452
F3	0.9696	-0.1316	1.1011
F5	0.8208	-0.3316	1.1524
F7	-0.5824	-1.4780	0.8957
FT7	-0.7943	-0.9024	0.1081
FC5	0.9581	0.0735	0.8846
FC3	1.8245	0.2906	1.5339
FC1	1.3696	0.0546	1.3150
C1	2.3185	1.5385	0.7800
C3	2.5526	1.6815	0.8711
C5	1.1109	0.6221	0.4888
T7	-0.8154	-0.8495	0.0341
TP7	-1.2328	-0.3816	-0.8512
CP5	1.2433	1.6590	-0.4156
CP3	1.9763	1.9028	0.0734
CP1	2.6456	2.3148	0.3309
P1	1.9774	2.4587	-0.4813
P3	1.0676	1.9210	-0.8534
P5	-0.0886	1.0972	-1.1858
P7	-1.8957	-0.0020	-1.8936
P9	-4.0307	-1.9745	-2.0562
PO7	-3.7790	-0.8413	-2.9377
PO3	-0.0290	1.3613	-1.3903
O1	-4.7467	-2.3117	-2.4350
Iz	-7.1480	-4.9873	-2.1607
Oz	-5.0913	-2.9397	-2.1515
POz	-0.5115	0.7093	-1.2208
Pz	2.8579	2.7757	0.0822
CPz	3.2673	2.5155	0.7518
Fpz	-0.1014	-0.8330	0.7316
Fp2	-0.0110	-0.8092	0.7983
AF8	-0.6439	-1.2307	0.5868
AF4	0.5044	-0.7042	1.2086
AFz	-0.4404	-1.5238	1.0833
Fz	0.0590	-1.2488	1.3078
F2	0.3633	-0.9586	1.3219
F4	0.9800	-0.2640	1.2440
F6	0.7293	-0.3874	1.1167
F8	0.3710	-0.6313	1.0023
FT8	-0.1631	-0.3921	0.2289
FC6	1.2064	0.1946	1.0118
FC4	1.8720	0.7614	1.1105
FC2	1.6869	0.1422	1.5447
FCz	0.5852	-0.7246	1.3098
Cz	2.2590	1.3079	0.9511
C2	3.0301	1.6950	1.3351
C4	2.0701	1.2078	0.8623
C6	1.6584	0.9192	0.7392
T8	0.3043	0.2833	0.0210
TP8	-0.3864	-0.0031	-0.3833
CP6	0.7774	0.4311	0.3463
CP4	2.4266	1.9234	0.5032
CP2	3.4484	2.6936	0.7548
P2	2.6522	2.9041	-0.2519
P4	1.0158	1.6687	-0.6529
P6	-1.0828	0.5089	-1.5917
P8	-2.6139	-0.5783	-2.0356
P10	-5.5844	-3.4183	-2.1661
PO8	-4.8141	-1.7568	-3.0573
PO4	-1.6018	-0.0687	-1.5331
O2	-5.6523	-2.7918	-2.8604

Table 1

3) Stats extractor 2020.m

In this example, we have two SPNs, one for anti-symmetry and one for symmetry. We can see them in ERP waves from electrodes PO7/8, and in the topoplots in Figure 7. Now we need to determine whether these SPN effects were statistically significant. We can test whether amplitude of the difference wave is significantly below zero, and whether symmetry is more negative than anti-symmetry.

To do this, we need amplitudes for each participant. This can be obtained with Stats extractor 2020.m outside the Grand Averages folder.

As with Waves 2020 and Topo 2020, you will need to change part of the script which sets the current directory, either to an equivalent place on your computer, or just comment these lines.

```
% DIRECTORY (THIS WILL NEED TO BE CHANGED TO A DIRECTORY ON YOUR
COMPUTER
folder1 = '/Users/alexismakin/Desktop/Alexis Work/Symmetry EEG/SPN
effect size and power/Project Folders/Project 8/EX5 AS4F';

folder2 = '/Users/alexismakin/Desktop/Alexis Work/Symmetry EEG/SPN
effect size and power/Project Folders/Project 8/EX5 AS4F/Grand
Averages';
cd(folder1)
```

The crucial lines are:

```
%TIME WINDOWS
low = 300;
high = 1000;
```

If you want to look at amplitude in the 300-1000 ms time window, leave these parameters as set. This will give amplitude for each participants (N = 22) and condition (Random, Anti-symmetry and symmetry) averaged across the 300-1000 ms window.

Line 21 shows the default electrodes for this experiment [25 and 62], which are PO7 and PO8. These are the default electrodes because they were used in the published paper.

When you run the Stats extractor 2020.m script, it will pause ask which ‘dogma’ you want to use in the command window (do not mistake this for the script freezing). As described in the paper, we tried running the same analysis using several different electrode clusters, which could have been used ‘dogmatically’ in all SPN research. For now, enter 0 when prompted to select a dogma, and use the default electrodes PO7 and PO8.

This script saves a ‘Results.mat’ file in the Grand Averages folder, and reopens it into the workspace. The three columns are three conditions (Random, Anti-Symmetry and

Symmetry) and the rows are participants. This format is suited for analysis in SPSS. Data is shown in first three columns of Table 2.

The order of the columns is determined by line 12.

```
conditionNames = {'Random','AntiSymmetry','Symmetry'};
```

If these entered in a different sequence, then the columns would be reordered.

Random	AntiSymmetry	Symmetry		AntiSymmetry	Symmetry
-1.650	-4.749	-5.932		-3.099	-4.282
-0.905	-3.818	-2.859		-2.913	-1.954
-0.895	-2.312	-3.848		-1.417	-2.953
-2.326	-4.994	-4.696		-2.668	-2.369
-0.137	-2.221	-3.727		-2.084	-3.590
-0.915	-2.102	-1.631		-1.187	-0.716
-3.873	-6.474	-6.473		-2.601	-2.600
-1.022	-1.208	-2.827		-0.186	-1.805
2.759	-0.983	-1.362		-3.741	-4.121
-0.436	-5.171	-4.704		-4.736	-4.268
0.430	-2.281	-2.363		-2.711	-2.793
-5.537	-8.128	-10.117		-2.590	-4.580
1.883	-0.380	-0.538		-2.263	-2.421
5.330	1.363	-0.454		-3.967	-5.784
-0.938	-2.892	-3.864		-1.953	-2.925
-1.138	-2.530	-3.601		-1.392	-2.463
-0.780	-3.973	-5.277		-3.193	-4.498
-6.828	-11.201	-10.395		-4.373	-3.567
-3.863	-4.353	-7.610		-0.490	-3.747
-4.226	-6.189	-7.297		-1.963	-3.072
1.060	1.700	0.385		0.639	-0.675
-4.574	-5.180	-5.336		-0.606	-0.763
			Means	**-2.250**	**-2.998**

Table 2

One option is to paste this output into excel and do further calculations therein. Table 2 shows this difference from random calculated in the last two columns (red), and the mean differences below (red) .

To double check these means, look at Table 1 electrodes PO7 and PO8, and average – we get -2.998, which is the same value. This kind of double checking is a good habit during EEG analysis – numbers should often by identical +/- very tiny differences caused by number of decimal places considered.

You can use SPSS or alternative software package to run one sample t tests on the yellow data and confirm that both symmetry and anti-symmetry amplitude is significantly below 0 ($p < 0.001$) and that symmetry generated a larger SPN than anti-symmetry ($p = 0.003$).

4) Thin stats extractor 2020.m

Some stats software prefers the data in a 'Thin' format, where conditions appear in different rows rather than as separate columns. The script 'Thin stats extractor 2020.m' will produce required format.

Many multi-level models need to code unbalanced aspects of the design, such as participant number and experiment number as well as interesting IVs and DVs. This labelling is done for you in Thin stats extractor 2020.m.

Furthermore, the SPNs are computed as difference waves (unlike in Stats extractor, which leaves this stage up to the user).

Lines 71 to 78 are crucial. As we have seen, Stats extractor 2020.m outputs Random in column 1, Antisymmetry in column 2, and Symmetry in column 3. Here we compute the difference from random, that is, column 2-1, and column 3-1. This is coded in the SPNcond parameter {[2 1], [3 1]}.

We see next that participants were attending to regularity in this experiment, so AttendReg is set to 1. The W scores from the regular stimuli are entered next [0.875 0.875].

If someone were to devise a new perceptual goodness score, they might replace condW with their own numbers, and run the same codes again, while leaving the factor labelling structure in place.

```
SPNcond= {[2 1] [3 1]};
AttendReg = 1;
condW = [0.875 0.875];
Project = 8;
Group = 19;
condNos = [1 2];
SPNIDs =[48 49];
firstSubjectNo = 398;
```

5) Peak extreme stats extractor 2020.m

This script was designed for analysis of P1 amplitude. One approach is to use the Stats extractor 2020.m and chose a window which capture the peak of the grand average P1. However, P1 is likely to peak at slightly different times for different participants, so this is likely to extract data not centred on the peak for each subject. A better estimate of P1 amplitude comes from the Peak extreme stats extractor 2020.m. This finds the maximum within a window, rather than the peak across the window.

This part of the code identifies the boundaries within which to search for the low and high peaks.

```
% TIME WINDOWS
      load timeVector
      low = 100
      high = 200
      iip = find(timeVector >=low & timeVector <=high);


      low = 150
      high = 250
      iin = find(timeVector >=low & timeVector <=high);
```

When we run this, we get a positive peak value for each participant in anti-symmetry and symmetry conditions, and a negative peak value in anti-symmetry and symmetry conditions (Table 3). These values look comparable to the P1 and N1 peaks in Figure 2. Note the smooth function substantially flattens these peaks in Figure 4.

	P1 peak estimate			N1 peak estimate	
	Anti Symmetry	Symmetry		Anti Symmetry	Symmetry
	0.4338	-0.6163		-8.3510	-11.0996
	3.6765	3.7824		-9.5602	-11.3593
	0.8569	2.6689		-8.0319	-6.6386
	0.7930	0.6503		-5.5585	-5.6691
	7.6244	6.5111		-11.7870	-14.5555
	2.0612	3.1249		-3.7864	-2.8476
	8.3934	9.8125		-5.9136	-6.0769
	5.5178	4.7180		-8.1990	-9.9758
	7.1464	6.1614		-7.7654	-7.6897
	8.4710	8.4183		-12.1415	-11.4091
	5.3979	6.7652		-4.8043	-7.6313
	3.8343	5.2280		-9.1795	-9.8676
	5.1701	5.6471		-0.2704	-1.1615
	4.3012	5.5806		-6.4807	-6.0246
	3.4332	1.9005		-5.8774	-5.5669
	5.9509	5.4106		-7.2075	-9.0288
	6.6556	6.9266		-8.7770	-9.2438
	6.9135	4.5130		-15.6098	-17.0652
	3.3762	0.1197		-8.5792	-14.7392
	3.0207	0.7199		-13.5141	-14.7833
	5.8798	5.8324		-3.3627	-4.2959
	-1.3573	-1.7848		-6.8077	-6.0280
Means	**4.4341**	**4.1859**		**-7.7984**	**-8.7617**

Table 3

6) Thin peak extreme stats extractor

Thin peak extreme stats extractor 2020 allows us to collected data for multilevel analysis, with the same labelling of potentially un nested conditions.

7) SPN effect size and power spreadsheet

Outputs from the stats extractor are compiled in a spreadsheet called *SPN effect size and power V8.xlsx*. (this will be updated as new experiments are integrated into the catalogue).

Sheet 1 summarises all data, Sheet 2 summarises double checks, but each sheet beyond that corresponds to one of the 40 projects. This is the official repository of SPN amplitudes, as computed with different clusters. It is worth checking whether you can recreate the numbers using Stats extractor 2020.

Here difference from random is calculated, along with of effects size and 95% CI.

Perhaps the most useful information for readers is on sheet 1. This has 249 rows, 1 for each SPN.

Columns code project folder, subfolder, SPN ID (compare with topoplots in Figure 4 of the manuscript) Number of participants in Experiment, Number of participants in a condition, nickname (possibly less useful outside our lab), whether the paper is in the world or file drawer, Lead researcher on the project, Journal published, year published, name of the regular condition, name of the irregular comparison condition, original electrodes and original time windows

Columns **V** and **W** code whether people are attending to regularity (1 or 0) and the W score of the regular pattern. Column X notes some caveats and ambiguities about W calculation (probably not interesting for most users).

Starting on Column **AA**, we have amplitude information for each SPN. The first few rows are shown in Table 4. The same information from alternative electrode clusters, and from left and right hemispheres are in columns **AL** to **CS**.

SPN	SD	D	N	p SPN	t	p	S.E.M	95% CI	minus	plus
-2.503059985	1.75111741	-1.4294073	24	0.91666667	-7.003	0.000	0.35744534	0.73943203	-3.242492	-1.763628
-2.318019857	1.65040195	-1.4045184	24	1	-6.881	0.000	0.33688689	0.69690363	-3.0149235	-1.6211162
-1.683175889	0.89687396	-1.876714	12	0.91666667	-6.501	0.000	0.25890521	0.56984653	-2.2530224	-1.1133294
-1.634350786	1.85196251	-0.8824967	24	0.79166667	-4.323	0.000	0.37803027	0.78201518	-2.416366	-0.8523356
-1.005652934	1.33545191	-0.7530432	24	0.75	-3.689	0.001	0.27259798	0.56391189	-1.5695648	-0.441741
-1.045343976	1.58719021	-0.6586129	24	0.625	-3.227	0.004	0.32398385	0.67021165	-1.7155556	-0.3751323
-1.568808625	1.9304458	-0.8126665	24	0.83333333	-3.981	0.001	0.3940506	0.81515577	-2.3839644	-0.7536529
-0.587145861	1.78435235	-0.3290526	24	0.58333333	-1.612	0.121	0.3642294	0.75346592	-1.3406118	0.16632005
-1.03537102	1.52177168	-0.6803721	24	0.625	-3.333	0.003	0.31063034	0.64258782	-1.6779588	-0.3927832
-0.410135075	1.32738417	-0.3089799	20	0.55	-1.382	0.183	0.29681212	0.62123492	-1.03137	0.21109984
-0.513750877	1.02740699	-0.5000461	20	0.75	-2.236	0.038	0.22973519	0.48084127	-0.9945921	-0.0329096
-0.3735052	1.26126439	-0.2961355	20	0.75	-1.324	0.201	0.28202729	0.5902899	-0.9637951	0.2167847
-0.244234791	1.3520194	-0.1806444	20	0.7	-0.808	0.429	0.30232073	0.63276456	-0.8769993	0.38852977
-0.841229057	1.44725015	-0.5812603	24	0.70833333	-2.848	0.009	0.2954187	0.61112014	-1.4523492	-0.2301089
0.135042634	1.12449354	0.12009196	24	0.375	0.588	0.562	0.22953628	0.47483198	-0.3397893	0.60987461
-1.093574482	1.43829937	-0.7603247	24	0.83333333	-3.725	0.001	0.29359163	0.60734056	-1.700915	-0.4862339

Table 4

3) Analysis in eLife paper

Analysis in the eLife paper **Lessons from a catalogue of 6674 brain recordings** is for the sake of transparency in the spirit of open science. Once this is downloaded and unzipped , you will have a folder and subfolders, and the annotated screenshots in Figure 1 and 2 point to the relevant files. Once you know which to look at, the names are quite self-explanatory. The 3 spreadsheets relating to the horsemen stem from the SPN effect size and power V8 spreadsheet. Distribution of SPN amplitudes comes from here as well. The Erratum file lists all the minor errors we encountered when scrutinizing our data sets – this may (unfortunately) need expanding in future! Other things not annotated are related to compilation and double checking.

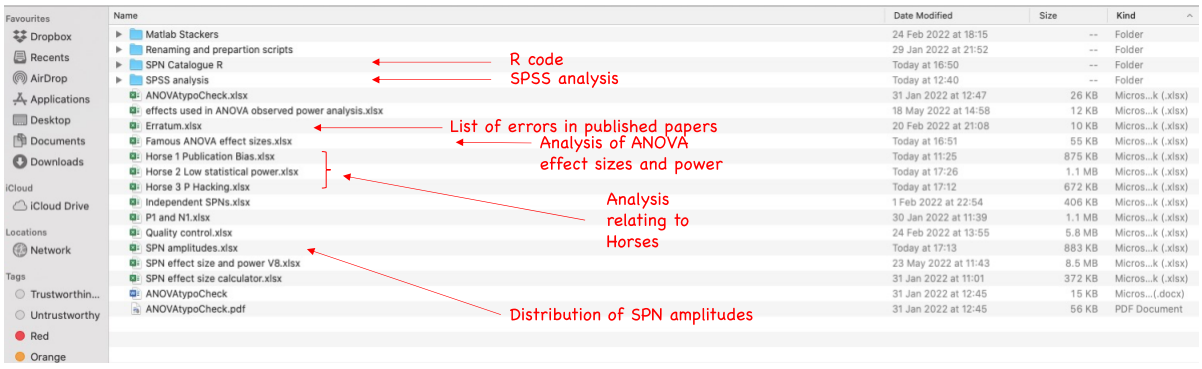


Figure 1

SPN Catalogue R is useful for the increasing number of researchers who routinely do analysis in R instead of Matlab Excel and SPSS. Once this is expanded, you can see the R code used for in various sections of the paper (Figure 2). The xlsx and csv files are mainly inputs for these r codes. No analysis is done within Excel here.

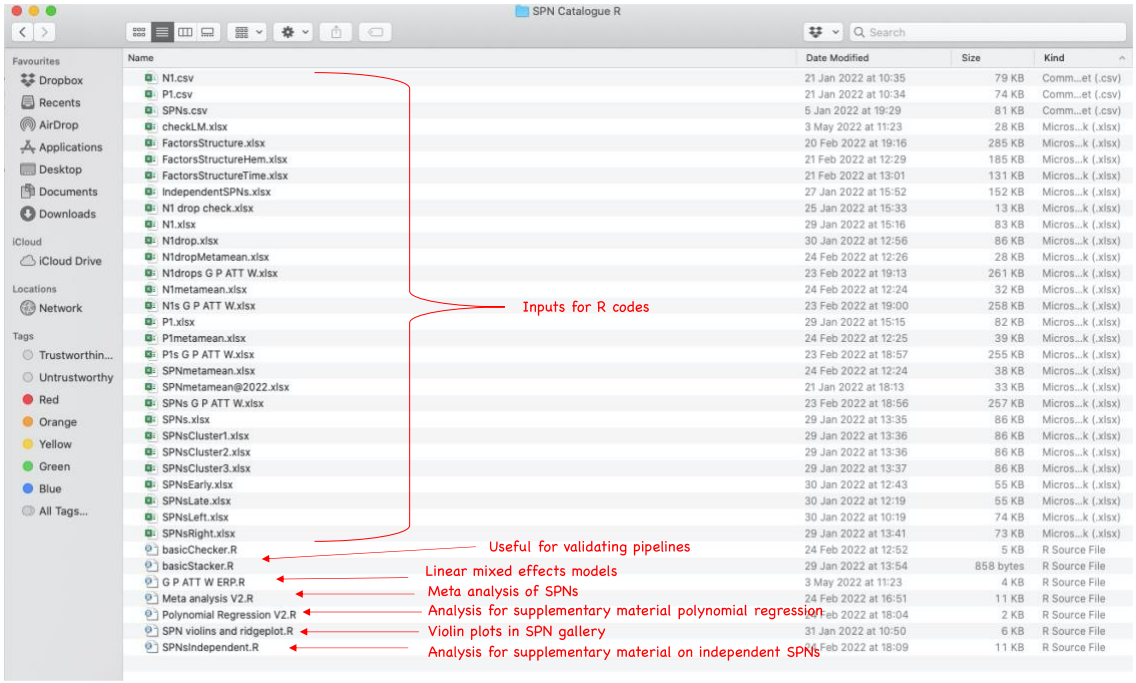


Figure 2

4) The SPN APP

Downloading App and Data

If you do not have MATLAB installed

1. This tool was developed using MATLAB. However, we have compiled this tool into a standalone executable file that only requires MATLAB Runtime to run. MATLAB Runtime does not require MATLAB to be installed. MATLAB Runtime is also royalty-free, and thus, it does not require a MATLAB license to be utilized. This app only requires that you download and install the appropriate MATLAB Runtime (if not installed already). First, download the repository from the following page (including dependencies and resources):

   https://github.com/JohnTyCa/The-SPN-Catalogue

   Next, head to the following website to download MATLAB Runtime 9.9.

   https://uk.mathworks.com/products/compiler/matlab-runtime.html

2. After installing MATLAB Runtime, the data is available to be downloaded on Open Science Framework at:

   https://osf.io/2sncj/

Note that the database comprises data from 40 projects requiring ~940 GB of storage, but you are free to only download the required projects - the tool will still work with a subset of the projects.

If you have MATLAB installed

1. For those who have MATLAB already, we have also made the app available in the form of a simple MATLAB function. Simply download the repository from the following page (including dependencies and resources):

   https://github.com/JohnTyCa/The-SPN-Catalogue

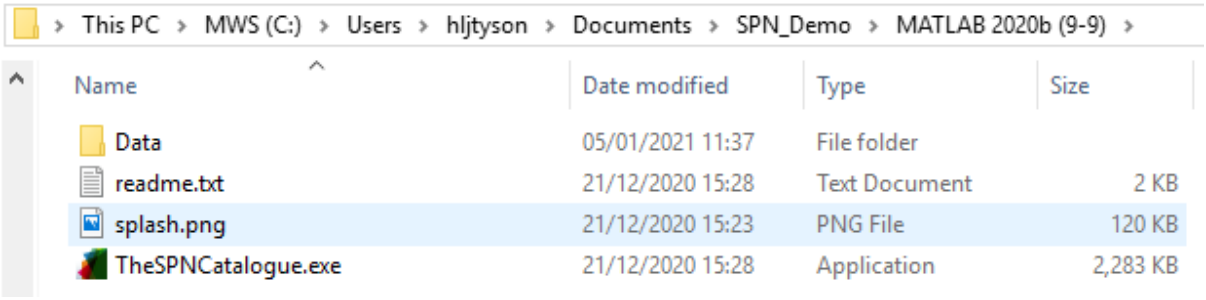
2. The required data is available to be downloaded on Open Science Framework at:

   https://osf.io/2sncj/

   Note that the database comprises data from 40 projects requiring ~940 GB of storage, but you are free to only download the required projects - the tool will still work with a subset of the projects.

Folder Layout

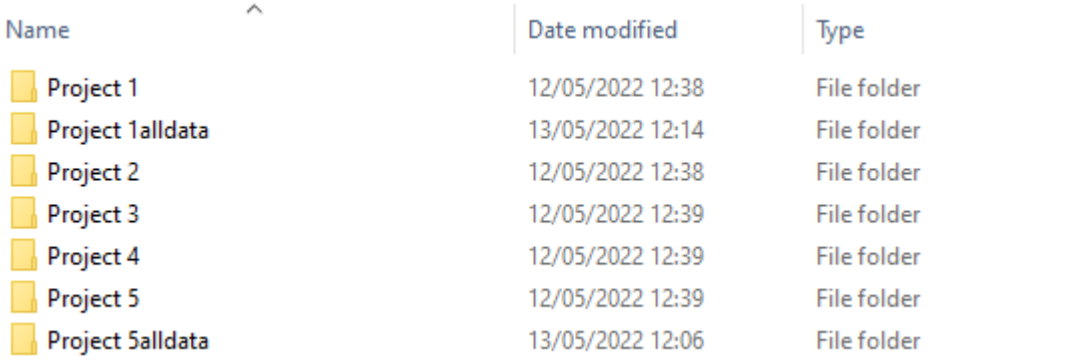
In order to run, the data needs to be organized in a certain manner. Below is an example of the folder downloaded from GitHub for MATALAB 2020b:



The contents of this folder include:

1) *Data* – This is the folder where the data will be located (obtained from OSF).
2) *readme.txt* – This is the readme file from GitHub.
3) *Splash.png* – The splash image for the app.
4) *TheSPNCatalogue.exe* – Executable app file.

As mentioned above, the “Data” folder will contain the data downloaded from OSF. Each folder within “Data” will correspond to data from a single project with the naming convention of “Project 1”, “Project 2”, “Project N”. See below for an example:



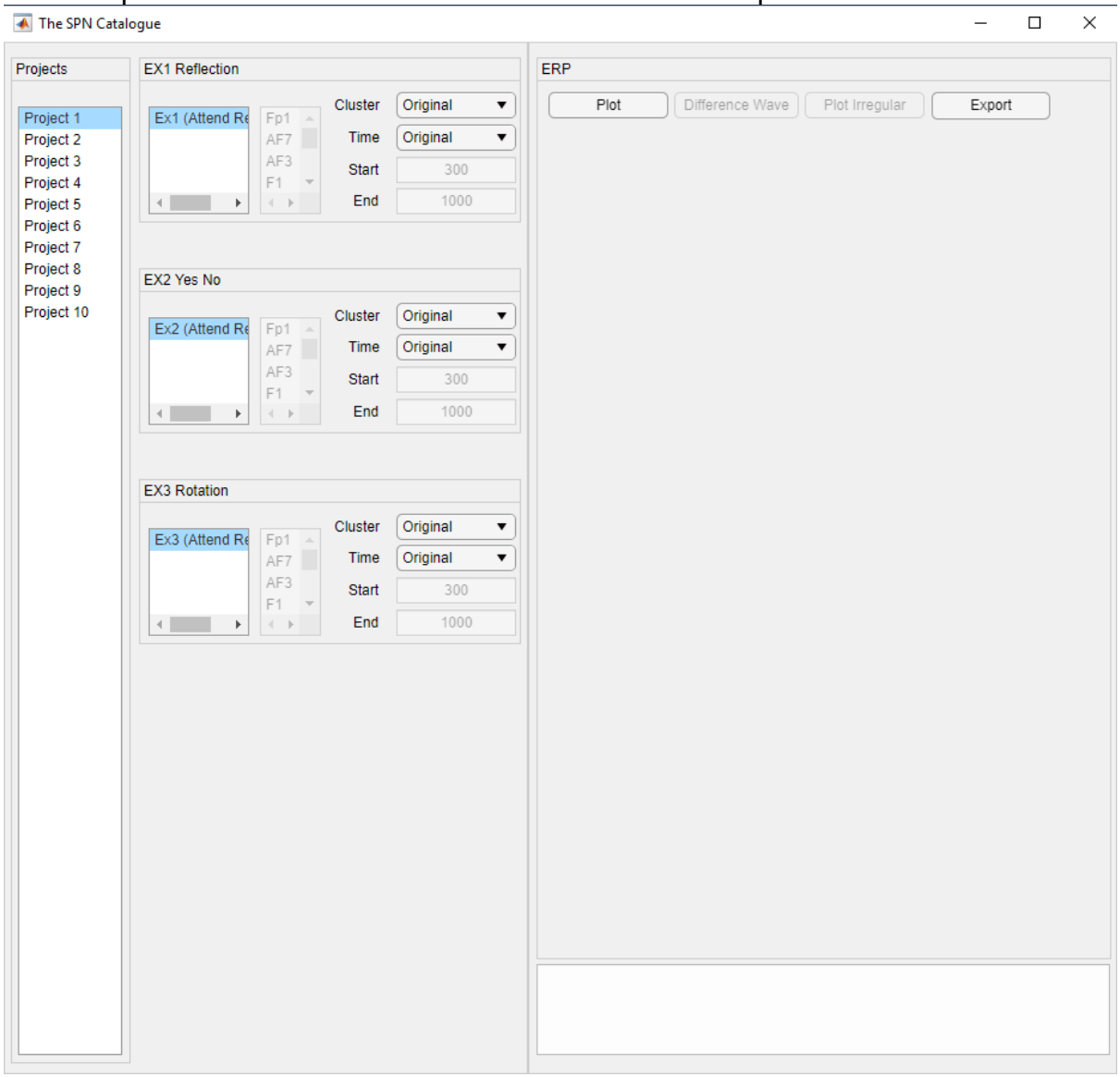
Note that above, only a subset of the 40 project folders are included since not all projects are required for the app to run. Therefore, you are free to download a subset of project folders. The catalogue also comprises “alldata” folders.

These folders are only required if you wish to export BIDS or SET files and are large in size, thus it is best to avoid downloading these unless required. You are able to only download a subset of the “alldata” folders if desired.

If the folders are organized like described above, the app should work fine.

Running the App

As long as the prior instructions have been followed, run the executable app file “TheSPNCatalogue.exe” in the root folder. If the “Data” folder is found in the same directory as the executable file, no further pathing is required. If the “Data” folder is located elsewhere on your device, you will be given the opportunity to select the location of the folder when the app initiates.
An example of what the window should look like when it opens is below:



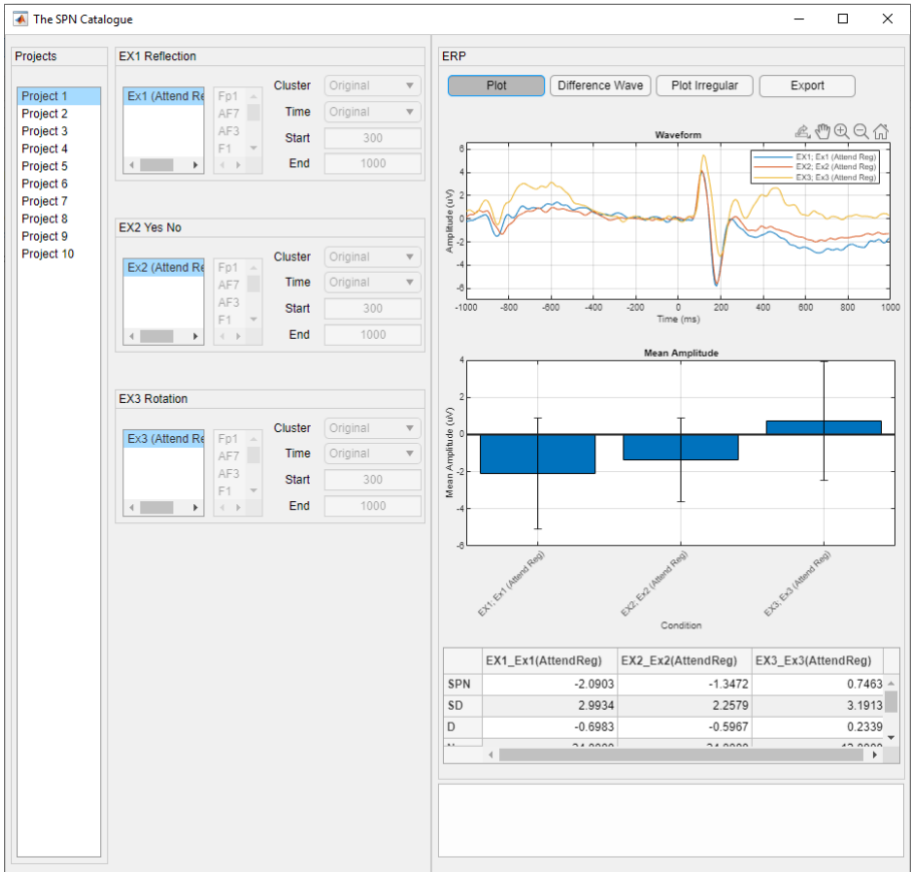
Visualizing Data

Once the app is open, you will see multiple projects listed on the left (corresponding to the project folders in the “Data” folder). Next to this list, you will see between 1 and 5 boxes corresponding to the individual experiments for the selected project. If the selected project is changed, the corresponding experiments will also refresh. For each experiment, you can see the conditions listed for that experiment.

5.1. Plot

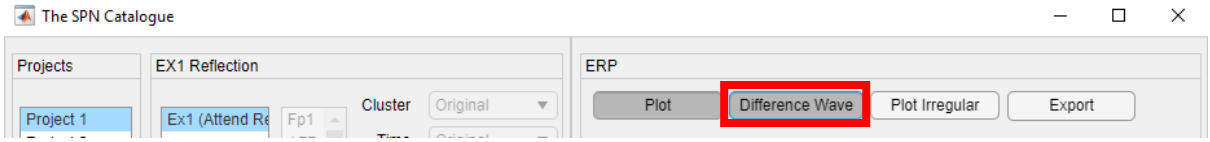
The “Plot” button allows for the plotting of data for all selected conditions. When “Plot” is clicked,

When “Plot” is clicked, the waveforms for the selected conditions will be displayed. You are free to select multiple conditions from the same experiment by holding down the “Ctrl” button on your keyboard and clicking the required conditions. On the right, you will also see the mean amplitude for the selected electrodes in the time interval selected for each experiment and condition. Below this bar graph, the summary statistics are shown in a table.

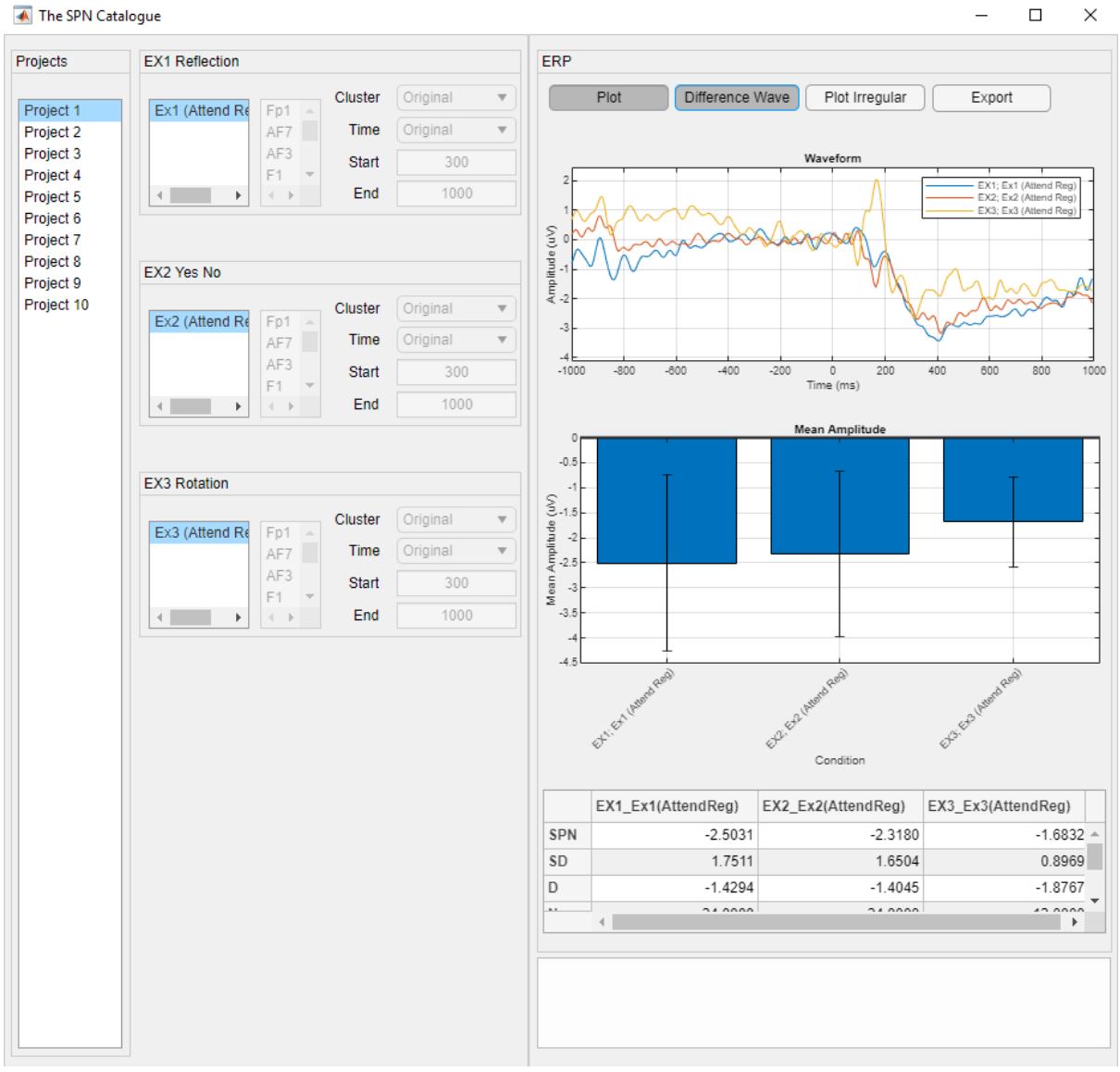


Difference Wave

The “Difference Wave” button allows for the visualization of the SPN wave, i.e. each condition has its corresponding “irregular” condition subtracted from it to produce regularity specific activity.

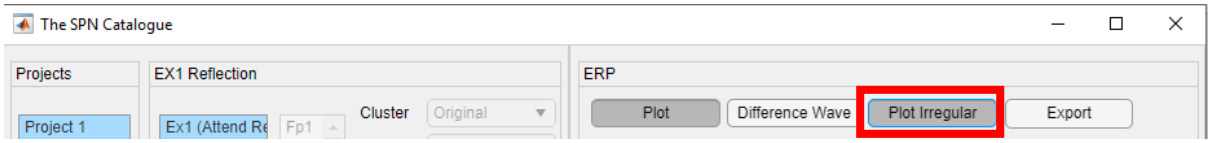


When “Difference Wave” is clicked, the waveforms, the mean values in the bar plot and the summary statistics in the table are all updated to represent the new values.

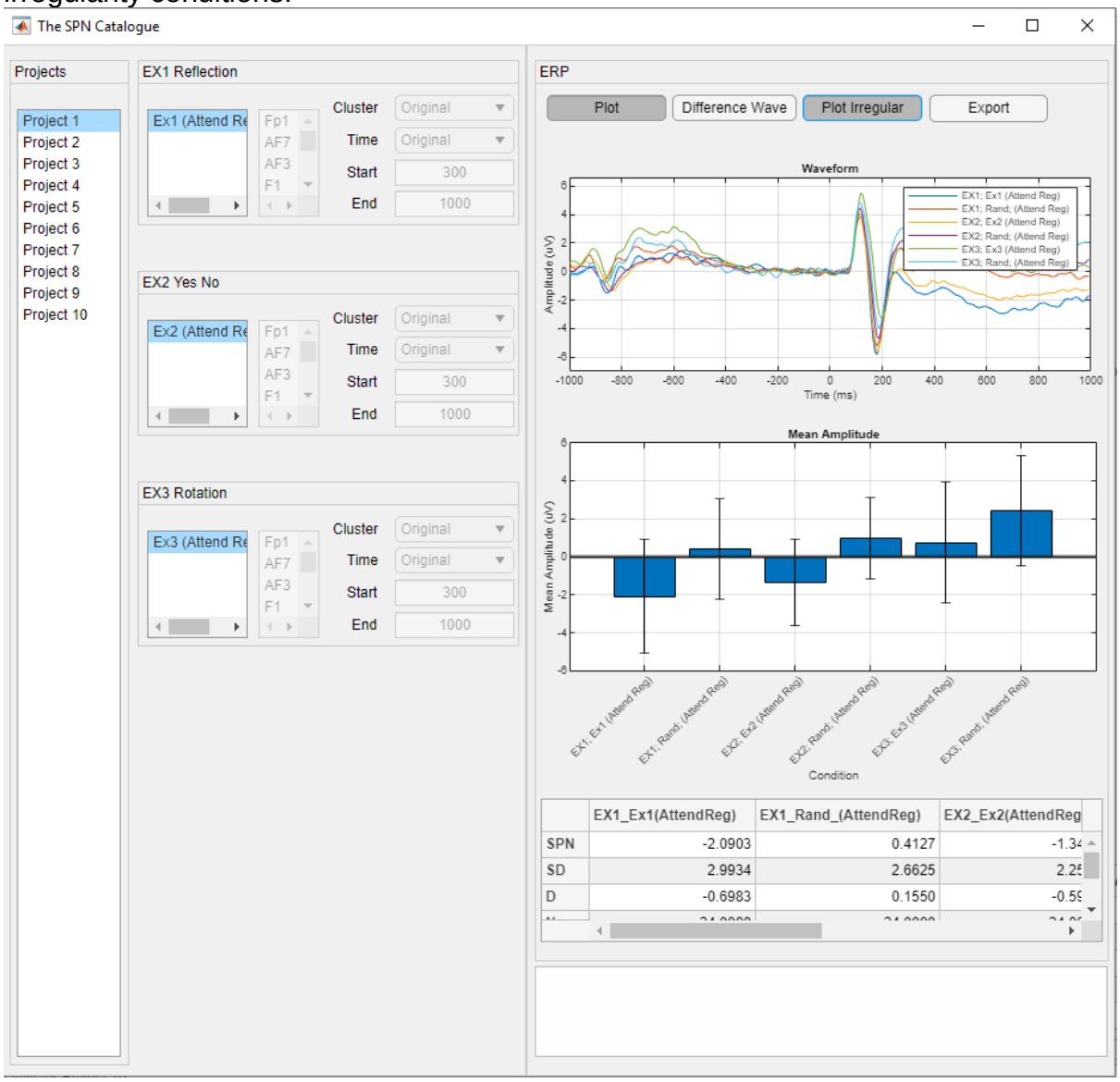


Plot Irregular

The “Plot Irregular” button allows for the plotting of the irregularity conditions for each experiment. Note that since the irregularity conditions are used to produce the difference waves, they will not be plotted whilst “Difference Wave” is also selected.

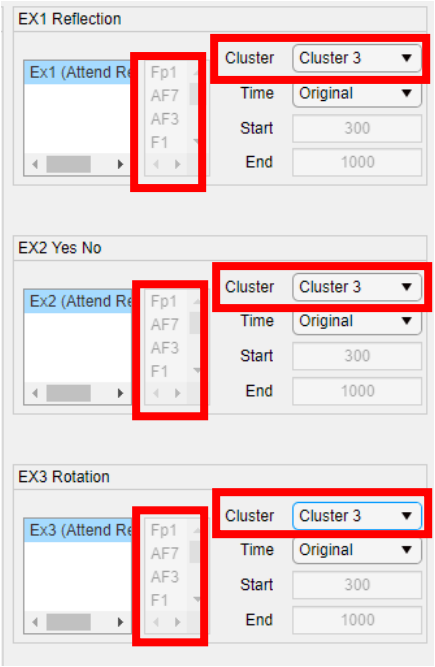


The waveform plot, the mean values bar graph and the table now include the irregularity conditions.



Cluster Selection

For each experiment, a cluster of electrodes were selected for the original analysis and reported within the corresponding manuscript. This app allows us to visualize the waveforms and mean values when different electrode clusters are selected.



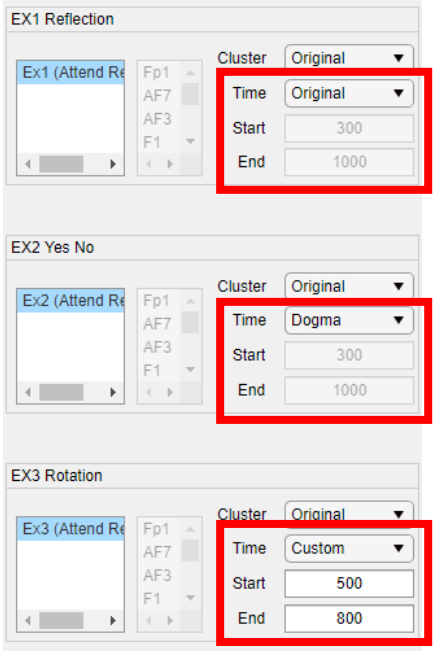
For each experiment, there is a dropdown box corresponding to the “Cluster” of electrodes. In addition to the “Original” cluster, other default selections are available under “Cluster 1”, “Cluster 2” and “Cluster 3”. Furthermore, you can select “Custom” in order to manually select a subset of electrodes from the list of electrodes. To select multiple electrodes, hold down “Ctrl” on your keyboard and click on the desired electrodes.

Please note that the selected electrode cluster is specific to the currently selected condition and the electrode cluster will need to be selected for each condition for each experiment.

Electrode clusters can only be changed when “Plot” is not selected. When the desired electrode clusters are selected, click “Plot” to update the values in the figures and table.

Time Selection

Similar to the cluster selection process, the time window can be selected for which to extract mean amplitude. Using the “Time” dropdown box, the “Original” time window is selected by default corresponding to the time window originally analyzed. To present alternative analysis windows, a second option of “Dogma” is available. This corresponds to an alternative time window that was not originally analyzed. Using the “Custom” option, a custom time window can be selected.

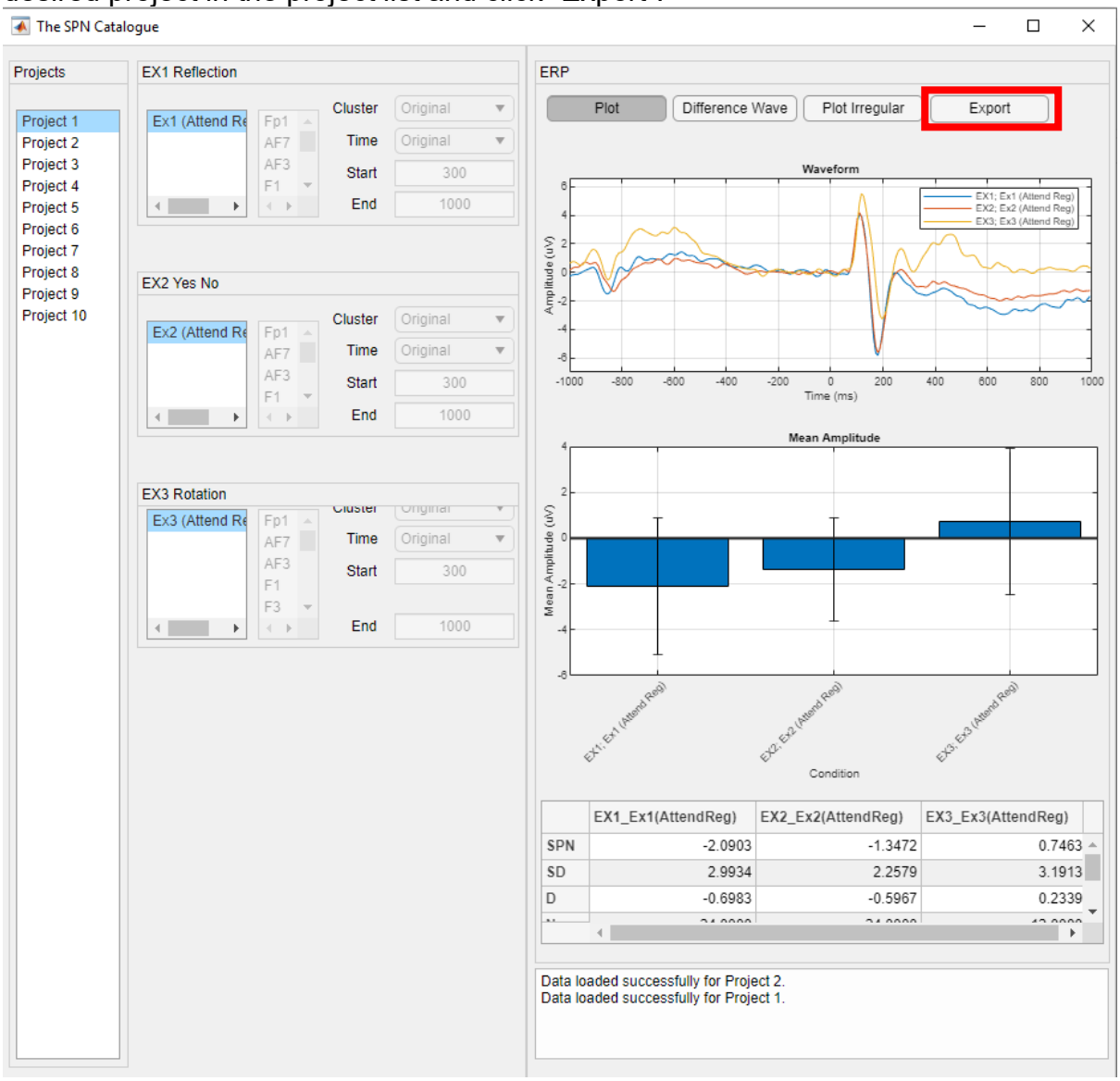


Please note that the selected time window is specific to the currently selected condition and the time window will need to be selected for each condition for each experiment.

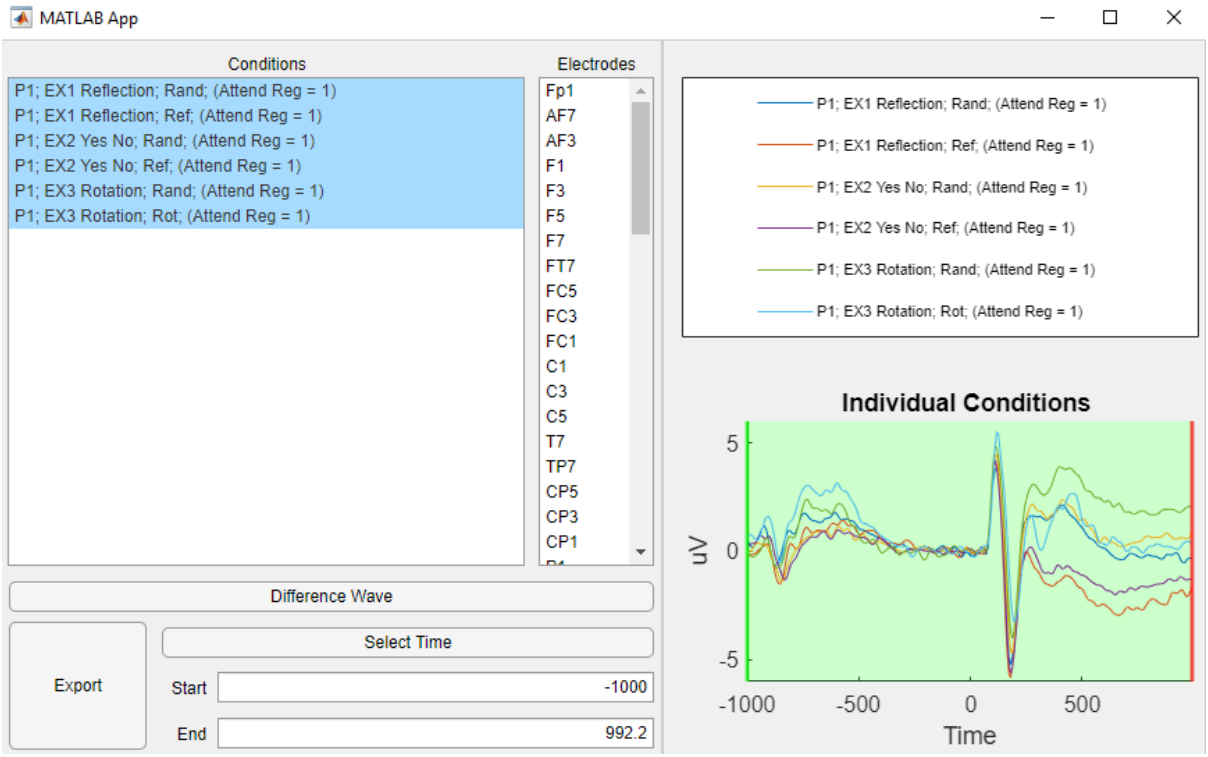
Time windows can only be changed when “Plot” is not selected. When the desired time window is selected, click “Plot” to update the values in the figures and table.

Exporting Data

In order to allow users to carry out further statistical analysis on the data, we have included an export function that will export the data to .csv format. Simply select the desired project in the project list and click “Export”.



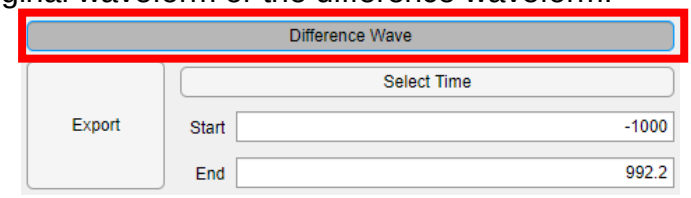
Below is the export window:



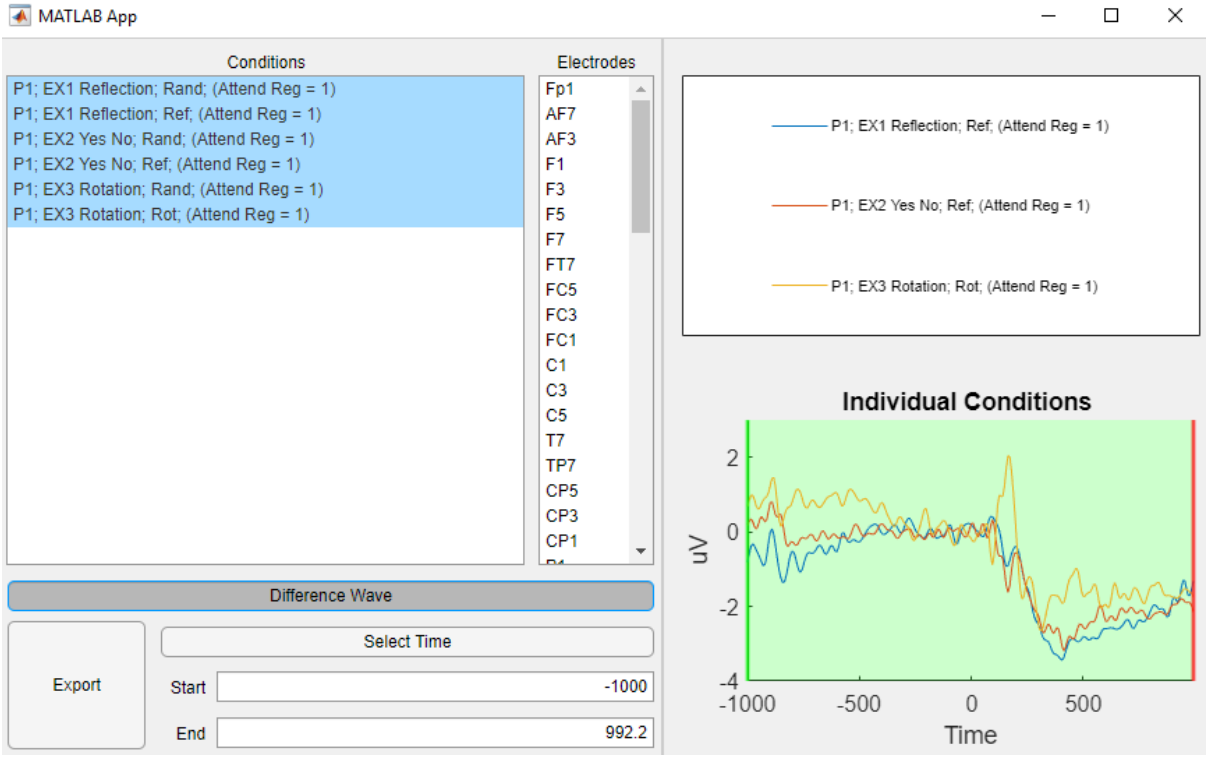
The conditions for the currently selected project will be listed on the right. The conditions for which we want to export data for can be selected. To select multiple conditions, hold “Ctrl” on your keyboard and click them.

Difference Wave

Similar to the main visualization window, you can export mean amplitudes obtained from either the original waveform or the difference waveform.

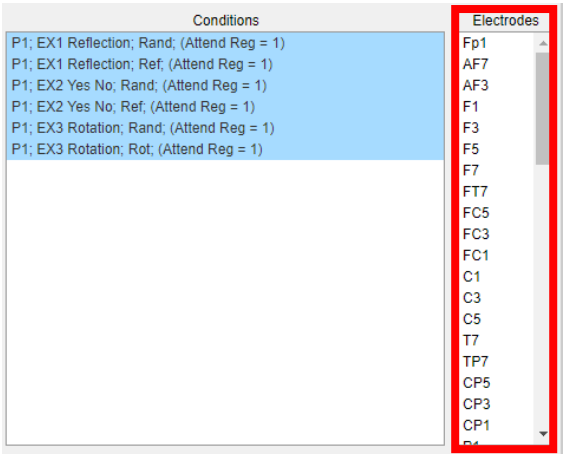


Once selected, the waveforms on the right will update to represent the difference from the corresponding irregularity condition.



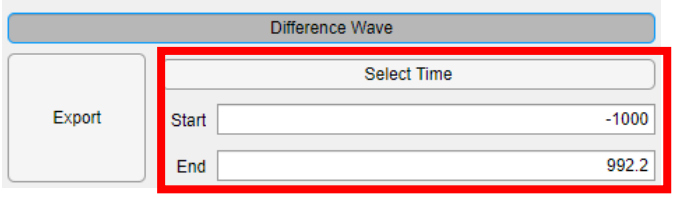
Cluster Selection

The “Electrodes” list allows us to select the electrodes for which we wish to extract mean values for. To select multiple electrodes, hold “Ctrl” on your keyboard and click the desired electrodes. The data across these electrodes will be averaged.



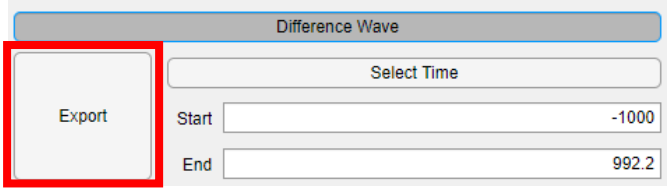
Time Selection

The time window we wish to average across can also be changed. The default selection will be the originally analyzed time window. To edit these values, either type in the desired “Start” and “End” time, or click “Select Time”. Clicking “Select Time” allows you to click on the waveform on the right to manually select the “Start” and “End” times.



Export

When the desired conditions, electrodes and time windows have been selected, clicking “Export” will allow you to select a folder to save the data to.



Once the desired folder has been selected, multiple files will be generated depending on the conditions selected:

1) A single “Long” format .csv file will be generated that will contain the mean amplitude for each subject and condition. The data from all experiments for that project will be contained within this file.
2) Multiple “Wide” format .csv files will be generated, one for each experiment in the project.

If the selected save folder already contains save data for the desired project, save names will be appended to avoid overwriting.

5) BIDS Datasets

Brain imaging Data Structure (BIDS) is a standardised way to organise neuroimaging data, allowing data to be organised consistently across any number of neuroimaging experiments regardless of lab and modality. The data for 78 experiments (from 84) present within the SPN Catalogue is available in BIDS format, each containing the raw BDF file from the original recording for each subject, as well as the derivative files resulting from the subsequent pre-processing and analysis. Where available, the code is also provided for each dataset. This data is available on OSF at https://osf.io/e8r95/.

Datasets are organised according to the project and the corresponding experiment number. For example, project 1 and its corresponding 3 experiments are sorted into “P001_EX1”, “P002_EX2” and “P003_EX3”.