BME3500 – Report

Report Type: Homework
Name and Surname: Toygar Tanyel
Student Number: 18011094
Date: 2 January 2023

Question 1:

a) To explore data files, following code is repeated and the results are shown:

```
load Q1_Data1.mat

% To figure out how we can reach inside of the data
```

whos command for Q1_Data1.mat

```
>> whos
  Name      Size            Bytes  Class     Attributes

  EEG       1x1           1933040  struct
```

for Q1_Data2

```
  Bytes  Class

1933840  struct
```

for Q1_Data3

```
  Bytes  Class

3476240  struct
```

- **All have the same structure, except for the "bytes".**

```
% whos provided "EEG" as name to explore data file
```

EEG for Q1_Data1.mat

```
EEG =

  struct with fields:

     srate: 500
      pnts: 1500
    trials: 40
    nbchan: 4
     times: [0 0.0020 0.0040 0.0060 … ]
      data: [4×1500×40 double]
```

for Q1_Data2.mat

```
EEG =

  struct with fields:

     srate: 400
      pnts: 1600
    trials: 30
    nbchan: 5
     times: [0 0.0025 0.0050 0.0075 … ]
      data: [5×1600×30 double]
```

for Q1_Data3.mat

```
EEG =

  struct with fields:

     srate: 600
      pnts: 2400
    trials: 60
    nbchan: 3
     times: [0 0.0017 0.0033 0.0050 … ]
      data: [3×2400×60 double]
```

The explanations of the parameters:

srate = sampling rate (for ex: for Q1_Data2 400Hz)

pnts = time points (for ex: for Q1_Data1, we have 1500 time points for each trials)

trials = number of trials (procedural repeat count for the experiment)

nbchan = number of channels (the amount of electrodes)

times = A vector of time points

data = A 3 dimensional structure with (nbchan, pnts, trials) which will be quite important in the other tasks.

b) For the first channel of all three datasets apply ERP analysis (use averaging to all trials for finding the ERP signal) both in time frequency domains. Please plot the average ERP signal in the time domain and the power spectrum of the average ERP signal in the same figure.

ERP basically helps us to avoid noises that randomly come from the brain oscillations. Therefore, it can be used to determine time-locked analysis because ERP will attenuate the random noises, whereas the time-locked activity will still remain as high.

```
% ERP is the time-domain average across all trials at each time point
erp = mean(EEG.data,3);
```

To analyze time domain, power spectrum and additional insights about the data1, data2 and data3, I utilized `plot_simEEG.m file` which is a code that was provided by Mike X Cohen[1]. I could also write the exact code from my old studies, however, this code is clean enough to analyze this kind of data for Question 1's requirements.
Integrated time domain and power spectrum codes that I wrote can be found in other questions.

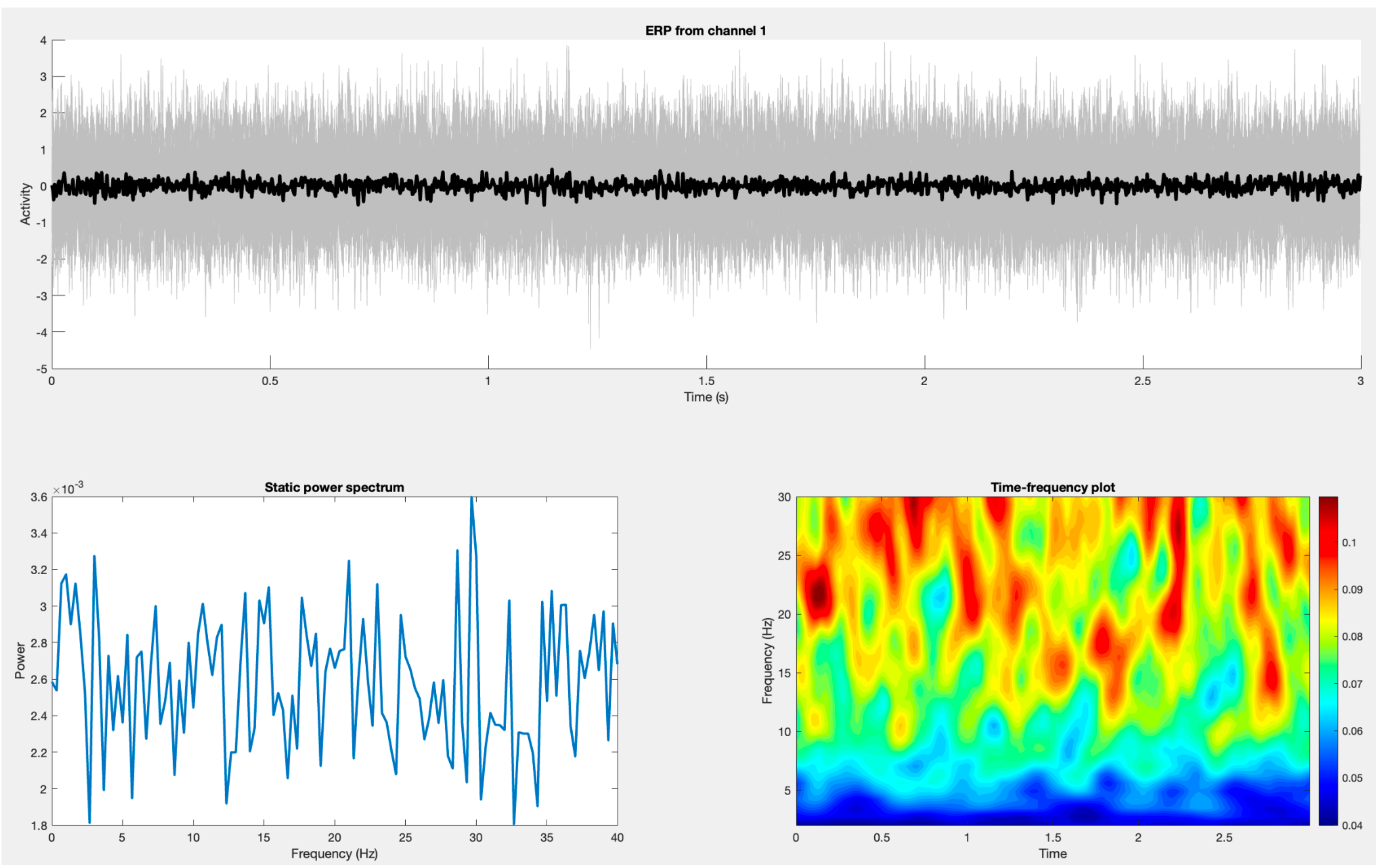


Figure 1: Time domain, power spectrum and time-frequency plots of Q1_Data1.m

[1] https://www.udemy.com/course/solved-challenges-ants/

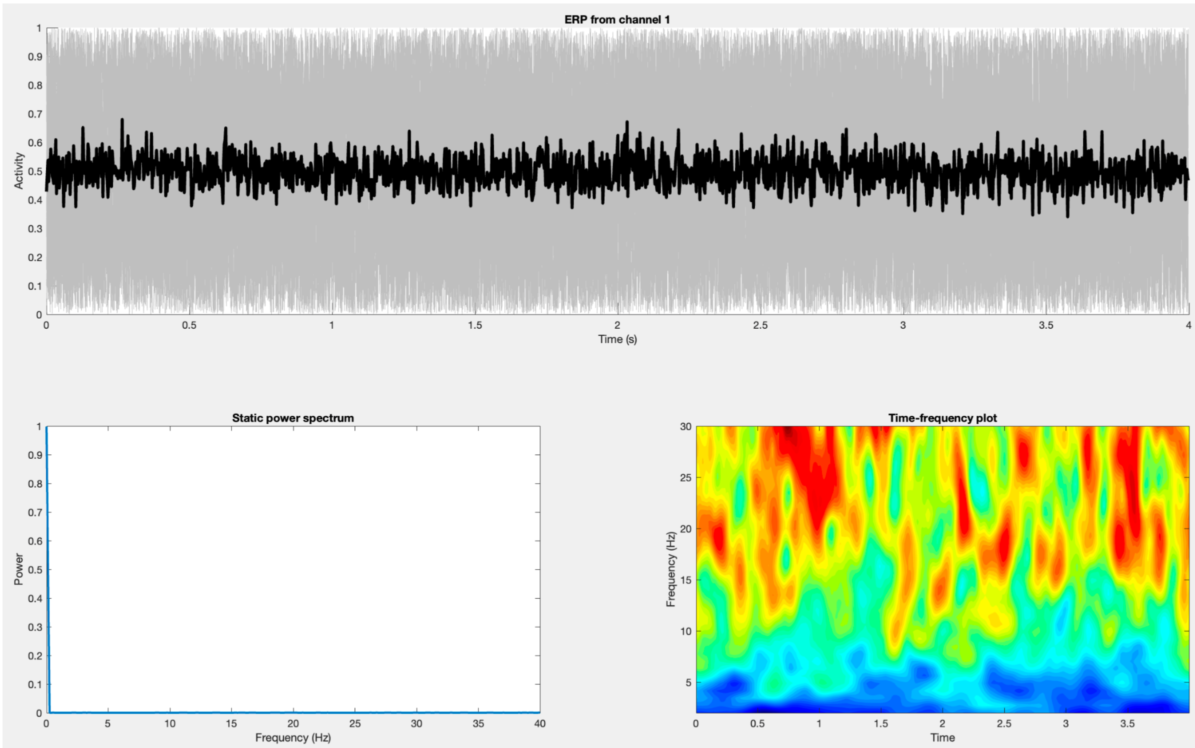


Figure 2: Time domain, power spectrum and time-frequency plots of Q1_Data2.m

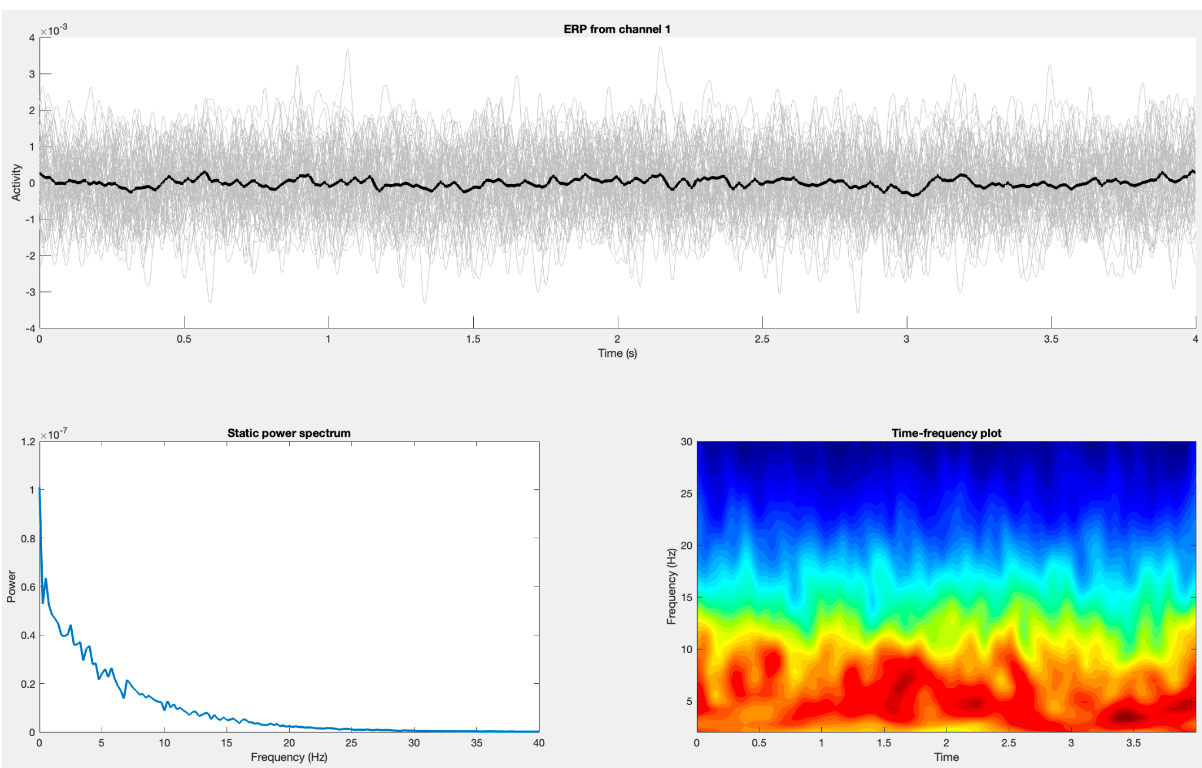
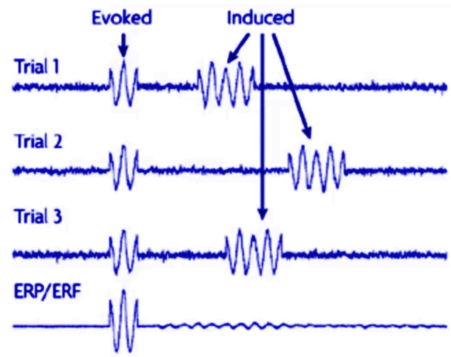


Figure 3: Time domain, power spectrum and time-frequency plots of Q1_Data3.m

Discussion about the figures of Data1, Data2 and Data3:

The faded gray signals in time-domain correspond to original voltage values before ERP and the black signal is the result of ERP. In every data, ERP decreased the voltage values as mathematically expected because we got means of the all trials. Also, variability in the single trial data is lost in the ERP averaging. Therefore, we lost some meaningful information in the single trials in attenuated values.



As we can see at the right illustration, we can use the "evoked" idea for time-locked analysis and the "induced" potentials that can be considered as non-phase locked activity are attenuated by ERP. I cannot extract meaningful interpretations by using peaks from our ERP results above for the task related potential, because in all 3 of the data the oscillations still look random which means that the data might be non-phase locked. In our results there are no obvious evoked potentials, and we can guess that these signals are artificial noises. On the other hand, we can see some of the differences between Data1, Data2 and Data3. There are almost no peaks in Data3 whereas Data1 and Data2 are quite frequent, but seem random. Data1 and Data2 have differences in their average voltage, and Data1 is between 0 to 3 seconds while Data2 and Data3 are obtained in 4 seconds. Therefore, their time points distributions might also be different.

Power spectrum provides some important information to interpret our data (discussed in "task c"). Time-frequency plots also provide additional information, although we cannot be sure without knowing the context of the task. The power spectrum and TF plot of Data1, Data2 and Data3 are explicitly different. The y axis of the static power spectrum corresponds to the amount of similarity or energy between the signal and the sine wave at different frequencies. The time-frequency plot can also be considered as a dynamic spectrum that shows how information changes over time.

c) **White Gaussian Noise, Uniformly Distributed noise and pink noise are three probability distributions which are employed while constructing the three datasets. If in each dataset only one kind of distribution was employed, please match each dataset to the proper noise distribution by using the finding that you have already obtained in section (b). Please add the necessary explanations.**

I did not realize c) is already providing an obvious clue about the data. As I mentioned above, ERP results were almost clear that we're dealing with non-stationary signals or artificial noises. However, the most indicative plot is the power spectrum. White, uniform and pink noise have very particular properties on frequency-power/amplitude domain. To create white nose, we can draw numbers from two different distributions, a normal (Gaussian) and uniform distribution. Uniform distribution of noise where all the values between zero and one are equally likely to be selected, and there is no possibility of getting numbers less than zero or greater than one. Pink noise is also hard to distinguish from white noise (Gaussian), and seems quite realistic. However, in the power spectrum plot, we can observe that pink noise has relatively low frequencies and has more power, and relatively higher frequencies have less power like Data3. From their structural differences, we can clearly see that Data2 is uniformly distributed and Data1 is White noise by looking at the power spectrum.
Summary: **Data1-White Noise**, **Data2-Uniformly Distributed Noise**, **Data3-Pink Noise**

Question 2: You are given a synthetic EEG data named "Q2_Data.mat". In this dataset, each trial is composed of the sum of three sinusoids having different frequencies and amplitudes. Additionally, there exists some degree of random noise in each separate trial.

a) Please plot the time domain averaged ERP signal for channel 1 and explain your findings. Can you make any guess about the amplitude or frequency values of three sinusoids from the time domain ERP signal? Give the necessary explanations.

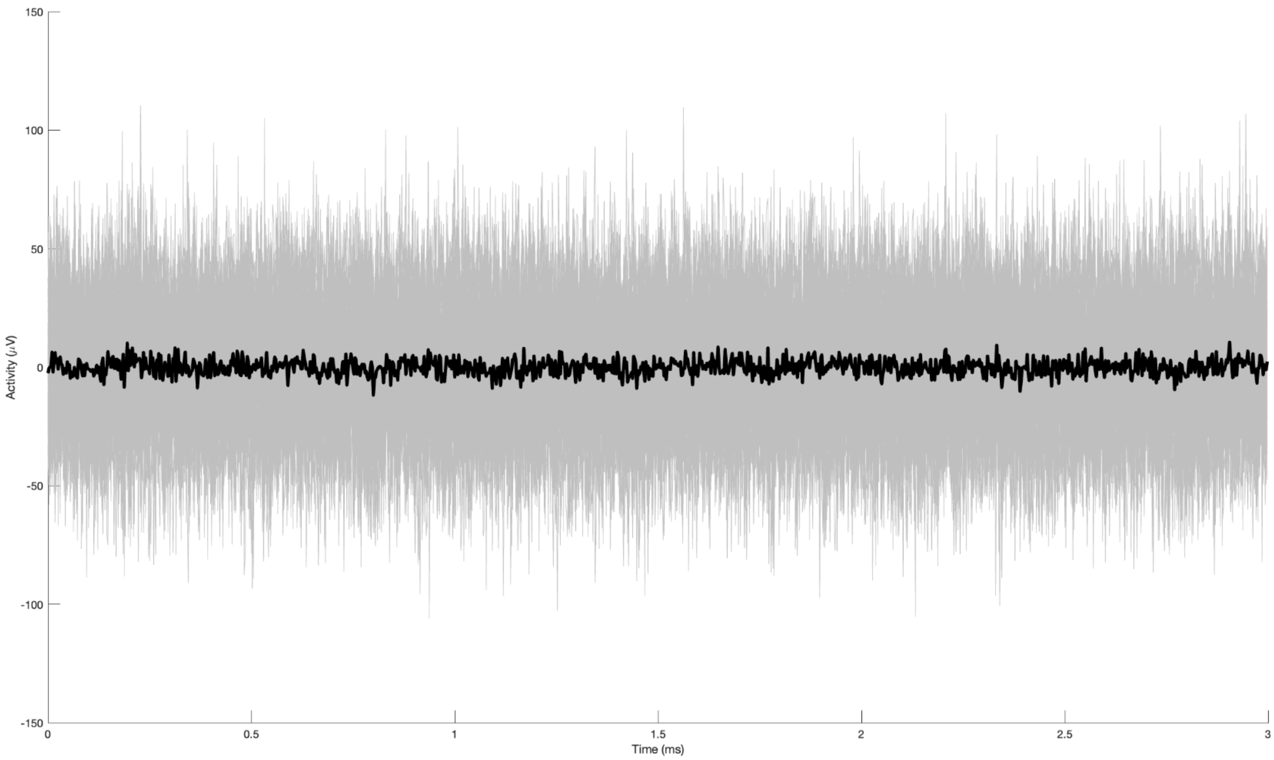
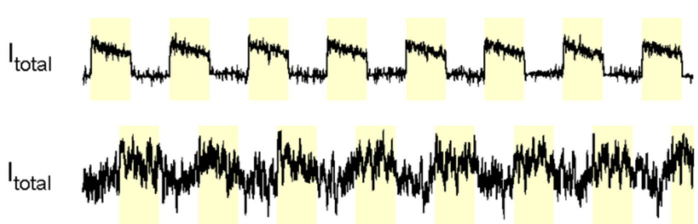


Figure 4: Question 2-a, Time domain averaged ERP signal for channel 1

We can barely see any patterns in the signal from the time domain. I guessed that noise might exist in the signal and provided figure[2] below. However, once again.., this information was already given in the question. Hence, my guess is still valid.

no noise:

noise added:



[2] Baden, T., James, B., Zimmermann, M. J. Y., Bartel, P., Grijseels, D., Euler, T., Lagnado, L., & Maravall, M. (2018). Spikeling: A low-cost hardware implementation of a spiking neuron for neuroscience teaching and outreach. In PLOS Biology (Vol. 16, Issue 10, p. e2006760). Public Library of Science (PLoS). https://doi.org/10.1371/journal.pbio.2006760

We would need to know more details about the signal and the three sinusoids in question in order to make a guess about the amplitude or frequency values of the sinusoids in a time domain. It is not possible for me to make a guess about the amplitude or frequency values of three sinusoids from a time domain ERP signal without any additional information or context. EEG signals, including event-related potentials (ERPs), are used to measure the electrical activity of the brain and are often analyzed in the time domain. However, the specific characteristics of an EEG signal, such as the amplitudes and frequencies of individual sinusoids, can vary widely depending on the experimental design, the type of EEG recording being made (e.g., scalp, intracranial), and the specific brain processes being measured. In order to accurately determine the amplitudes and frequencies of sinusoids present in an EEG signal, it is typically necessary to apply techniques from signal processing and spectral analysis, such as Fourier transforms or wavelet decomposition, to extract the relevant information. It is also important to consider the noise and artifacts present in the signal and to apply appropriate filtering or correction techniques to remove or reduce their influence.

b) Plot the power spectrum representations:

- **Power spectrum of the first trial of channel 1. Use the x-axis limits as [0 100] Hz.**
- **Power spectrum of the averaged trials in time domain for channel 1.**
- **Power spectrum calculated by averaging the Fourier representations of individual trials for channel 1.**

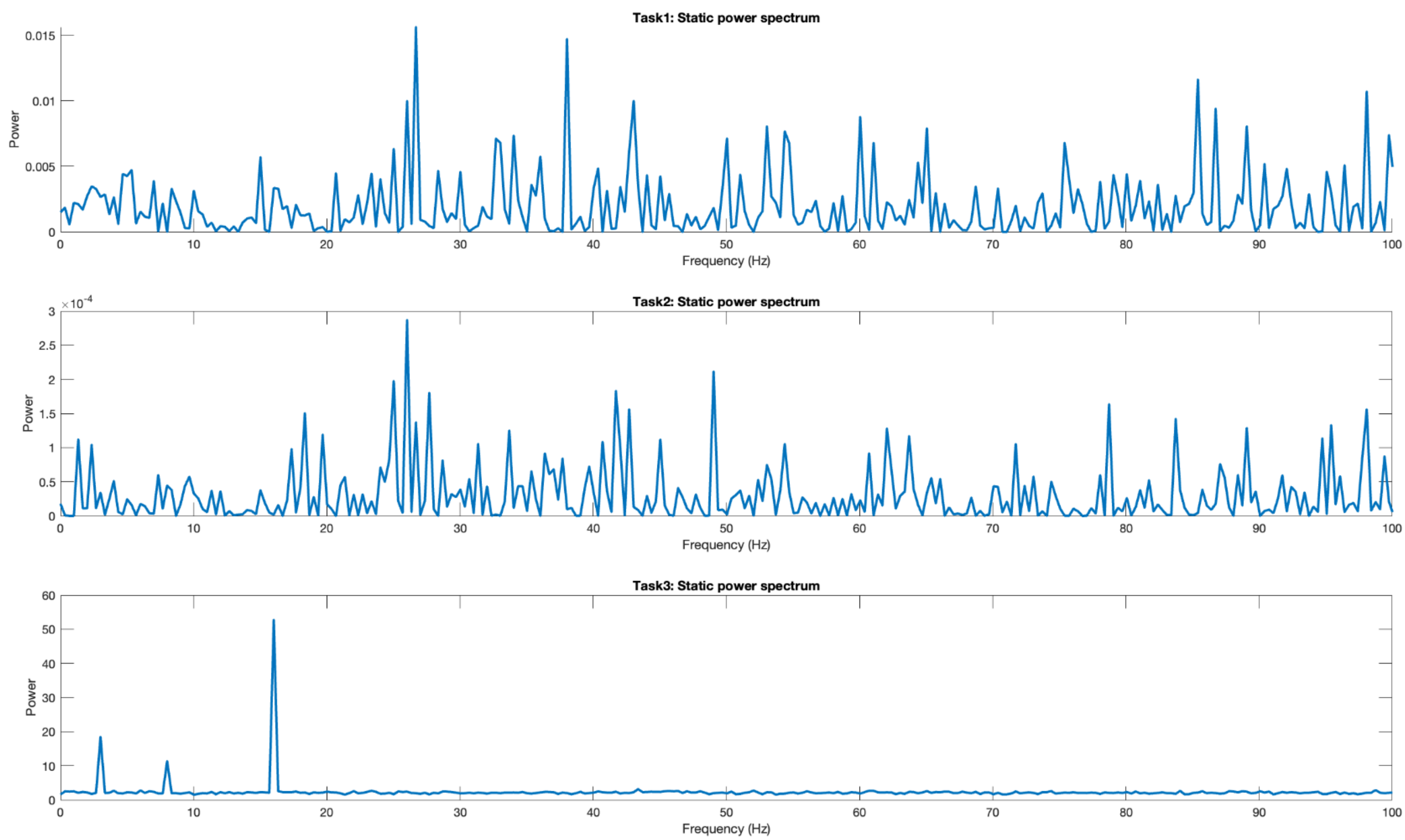


Figure 5: Different power spectrums for channel 1 of Q2_Data. Task1: first trial, Task2: averaged trials in time domain, and Task3: averaging the Fourier representations of individual trials.

The amplitude and frequency values of the sinusoidal components of a signal can be determined using the power spectrum of the signal. The power spectrum, which is a plot of the signal's power (or energy) as a function of frequency, can reveal details about the signal's frequency content and the relative strength of its different frequency components. We can utilize the **third** power spectrum of an EEG signal and look for peaks or "bumps" in the spectrum that correspond to the frequencies of the three sinusoids to determine the amplitude and frequency values of the sinusoids. The heights of the peaks in the power spectrum can be used to estimate the amplitudes of the sinusoids, and the locations of the peaks on the x-axis can be used to estimate the frequencies (i.e., the frequency axis). In the **third plot** we have 3 clean peaks at 3 hz, 8 hz and 16 hz with power of ~18.5, ~11.4 and ~52.7, respectively.

With the information given by the power spectrum, it is hard to tell whether the activity depicted on channel 1 is phase-locked or not. These activities are explained in Question 1. Knowing more about the experimental setup and the particular stimulus or event that the activity is being recorded in response to would be necessary to determine whether the activity represented in channel 1 is phase-locked or non-phase-locked.

Question 3: You're given a synthetic EEG data named "Q3_Data.mat". In this dataset, each trial is composed of the sum of two band-limited signals which have different center frequencies (can be interpreted as the peak values) and bandwidth values. Additionally, there exists some degree of random noise in each separate trial. ~ This task is almost the same with Question 2, only with the signal structure and some interpretations change ~

a) **Please plot the time domain averaged ERP signal for channel 1 and explain your findings. Can you make any guess about the center frequency or bandwidth values of these two band-limited signals from the time domain ERP signal? Give the necessary explanations.**

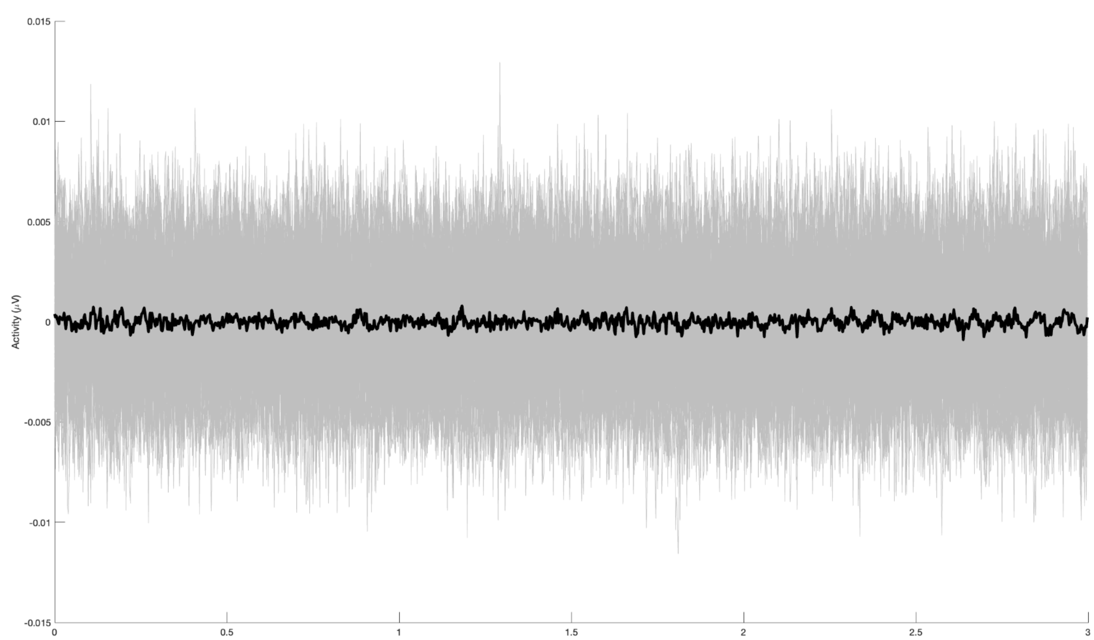


Figure 6: Question 3-a, Time domain averaged ERP signal for channel 1

It is generally challenging to infer a signal's center frequency or bandwidth solely from its time domain. This is so because a signal's frequency content cannot be determined from the time domain representation of the signal. A band-limited signal must be analyzed in the frequency domain in order to find its center frequency or bandwidth. Techniques like spectral analysis or filtering can be used for this. Spectral analysis involves computing the Fourier transform of the signal and examining the magnitude or power spectrum of the resulting frequency domain representation. The bandwidth of the signal can be calculated from the width of this peak, where the center frequency of a band-limited signal corresponds to a peak in the magnitude or power spectrum.

b) Plot the power spectrum representations calculated for following three scenarios on the same figure. Please explain which Power spectrum representation can be used to find center frequencies and bandwidth. Which band-limited activity has a wider frequency distribution?

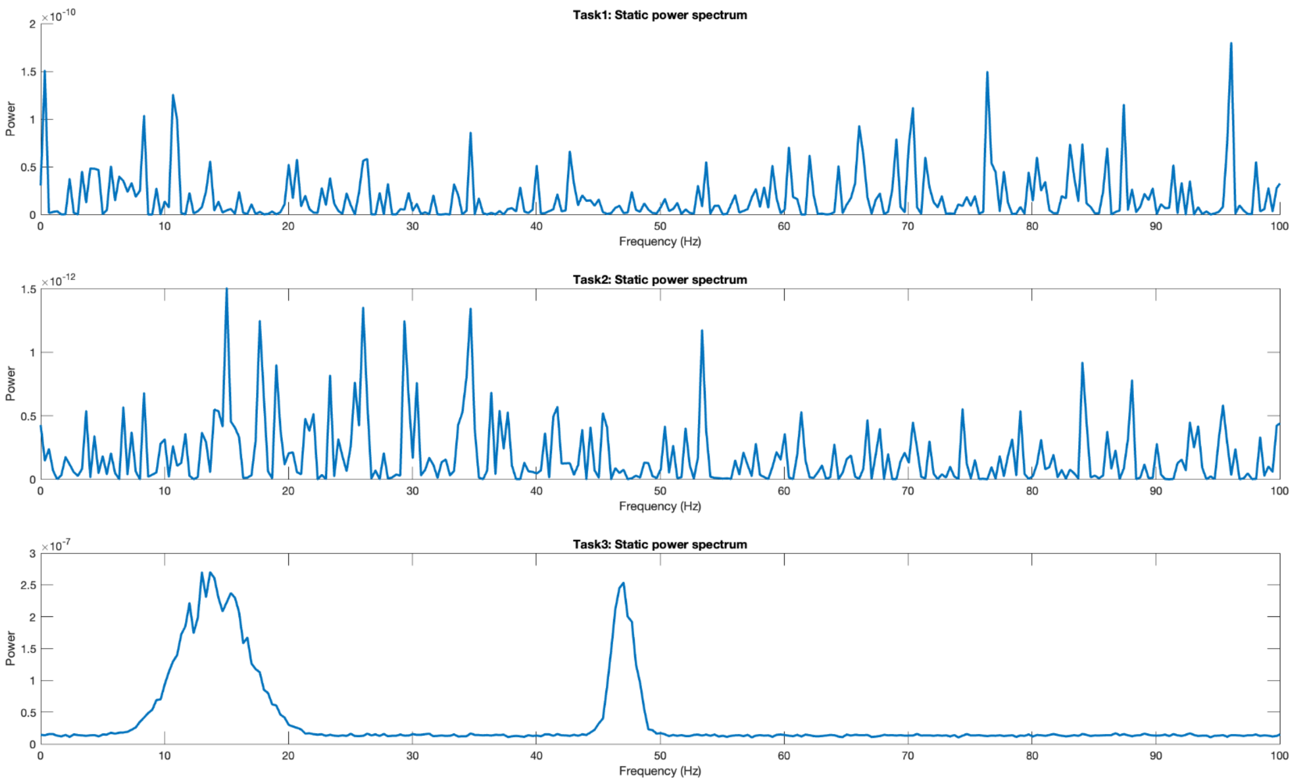


Figure 7: Different power spectrums for channel 1 of Q3_Data. Task1: first trial, Task2: averaged trials in time domain, and Task3: averaging the Fourier representations of individual trials.

To find the center frequencies and bandwidth of a signal using its power spectrum, we can look for the peaks in the power spectrum and measure their frequencies and the width of the peaks. The center frequency of a signal is the frequency at which the power spectrum is peaked, and the bandwidth is the range of frequencies over which the power is distributed. Hence, by looking at the **third plot,** we can make inference that the left-hand side signal has a wider range of frequencies. Ultimately, the signal that corresponds to the **~13.7hz center frequency** has a broader frequency range than the signal that corresponds to the **~47hz center frequency.**

Question 4: You're given a synthetic EEG data named "Q4_Data.mat". In this dataset, in each trial a band-limited signal, which can be called as a transient activity, is located onto a specific time point. Additionally, there exists some degree of random noise in each separate trial.

a) Please plot the time domain averaged ERP signal for channel 1 and explain your findings. Can you make any guess about the location and bandwidth of this transient activity from the time domain ERP signal? Give the necessary explanations.

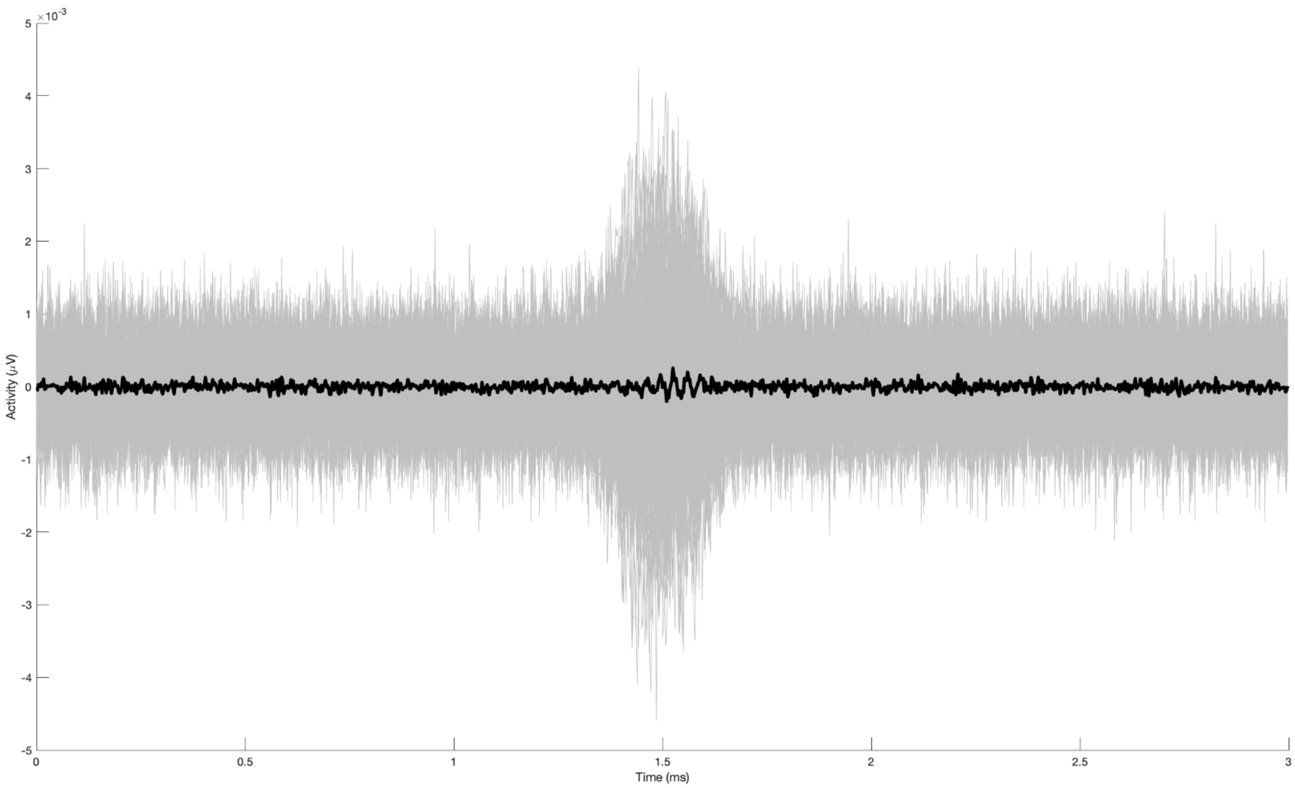


Figure 8: Question 4-a, Time domain averaged ERP signal for channel 1

By observing the form and size of the activity in relation to the surrounding noise, we can infer the location and bandwidth of the transient activity in the time domain ERP signal.

We can search for peaks or spikes in the signal that happen at a particular time point to determine the location of the transient activity. These peaks or spikes may indicate the presence of a transient activity. The bandwidth of the transient activity could potentially be determined by examining the duration of the deviation.

In order to accurately locate and identify the transient activity, it is also crucial to take into account how noise affects the signal. To increase the signal-to-noise ratio and make it simpler to spot transient activity, it may be necessary to use noise reduction techniques or average the signal across multiple trials if the noise is significant.

b) Plot the Short Time Fourier Transform representation of the time-averaged ERP activity by using the "spectrogram()" built-in function of MATLAB. Try different window types, window size, overlap-ratio and FFT size to be able to obtain the best representation. From the Spectrogram visualization, please try localizing transient activity in time and try to find the peak frequency and bandwidth of the same activity in frequency.

To determine the bandwidth of the transient activity, we can also analyze the frequency content of the signal. If the transient activity has a narrow bandwidth, it will be represented by a sharp peak in the frequency domain. If the transient activity has a wider bandwidth, it will be represented by a broader peak in the frequency domain.

```
% input parameters
window_sizes = [175, 256, 512];
overlap_ratios = [150, 175];
fft_sizes = [128, 256, 500, 512];
window_types = hann, kaiser, flattopwin
->colormaps: hot, bone, jet, respectively.
```

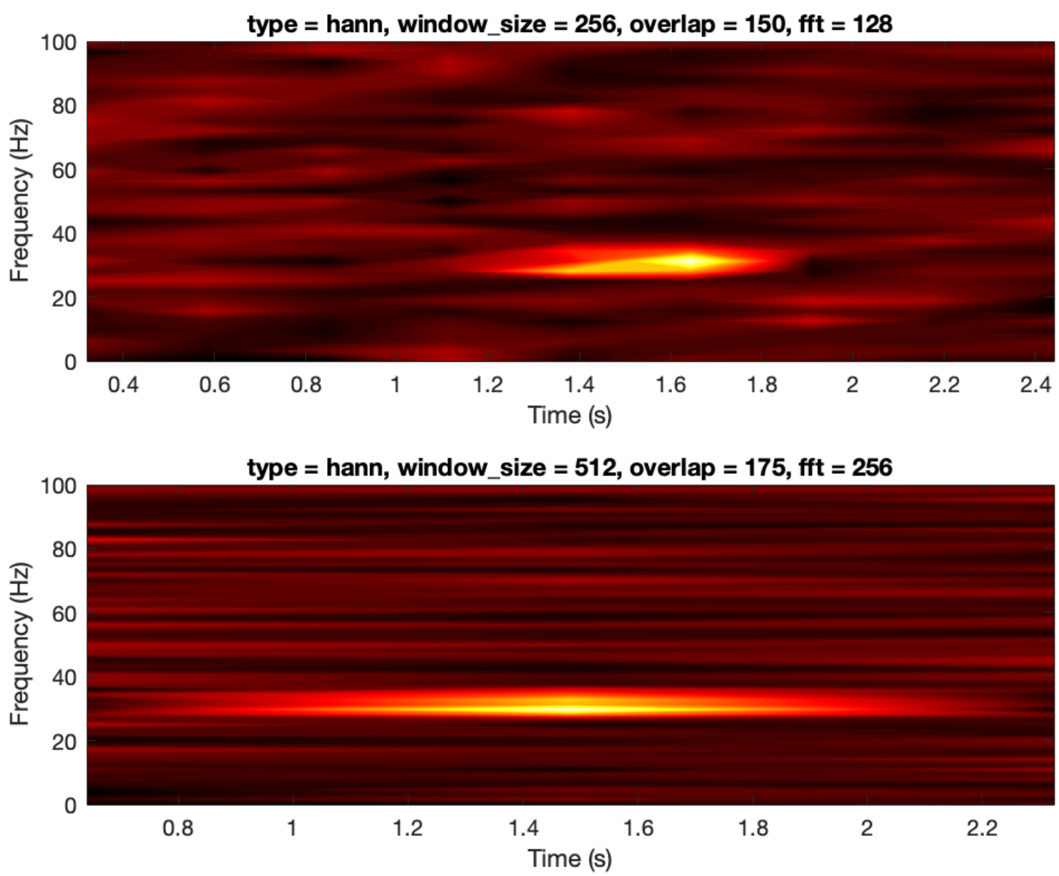


Figure 9: Test cases for hann.

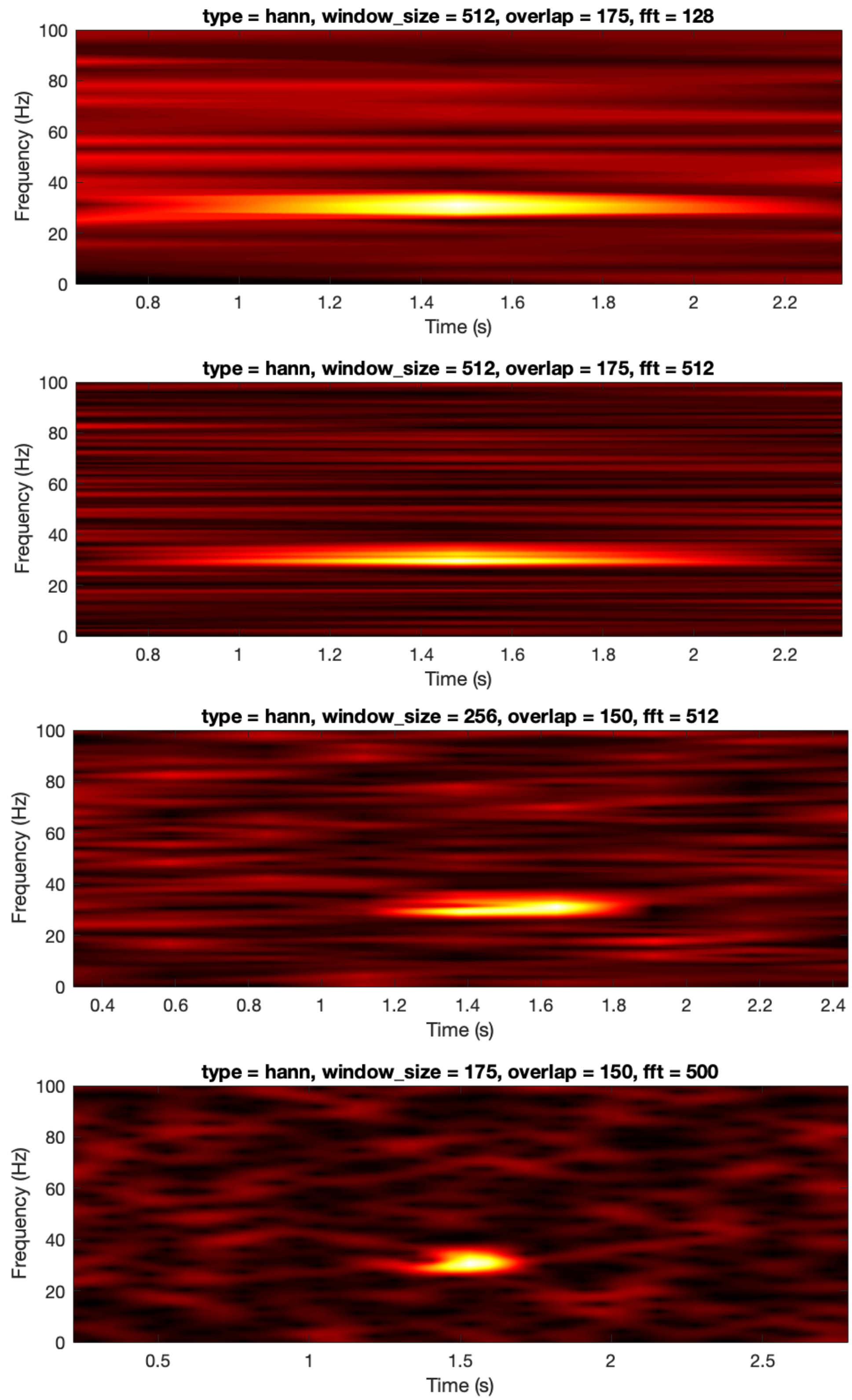


Figure 10: Test cases for hann.

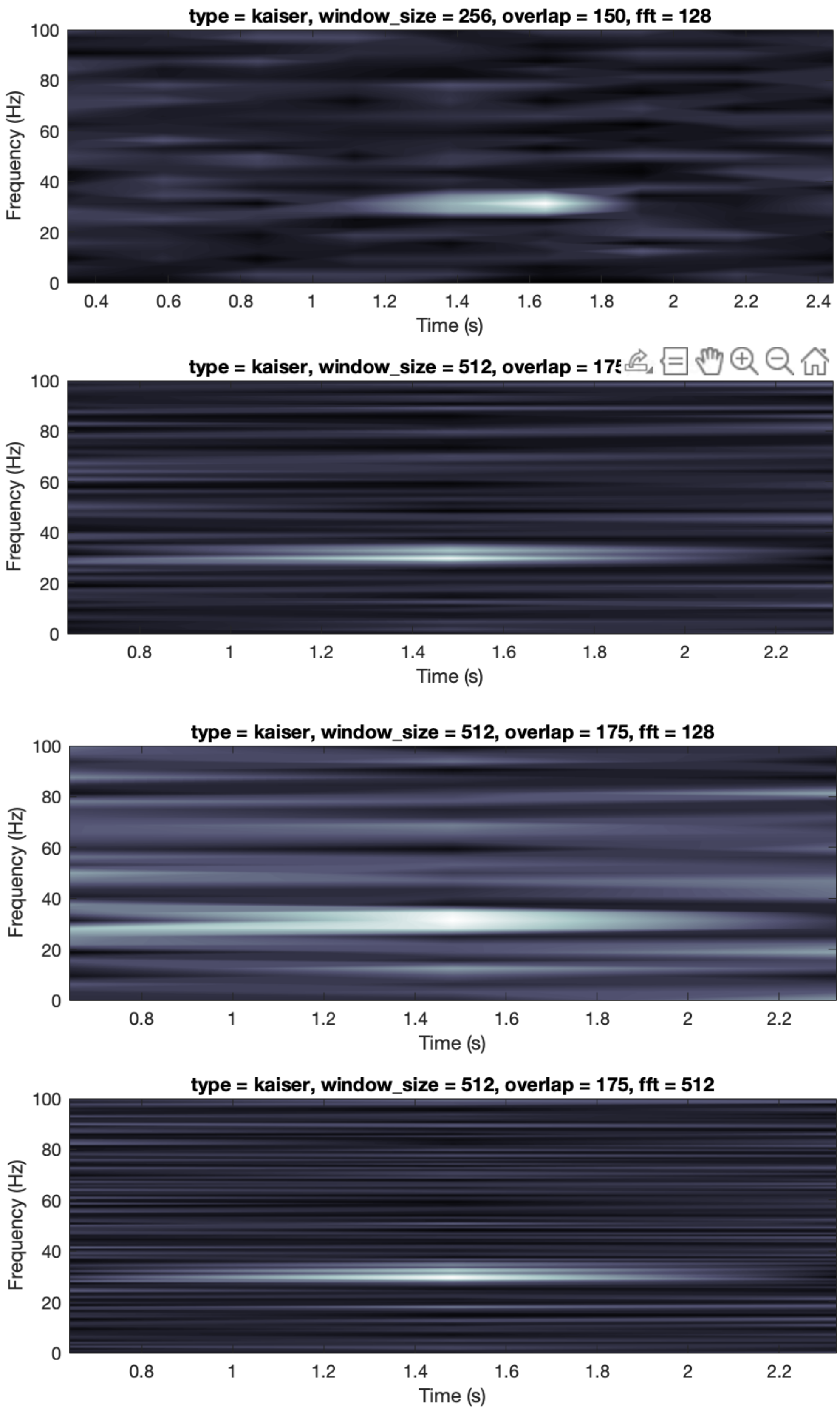


Figure 11: Test cases for kaiser.

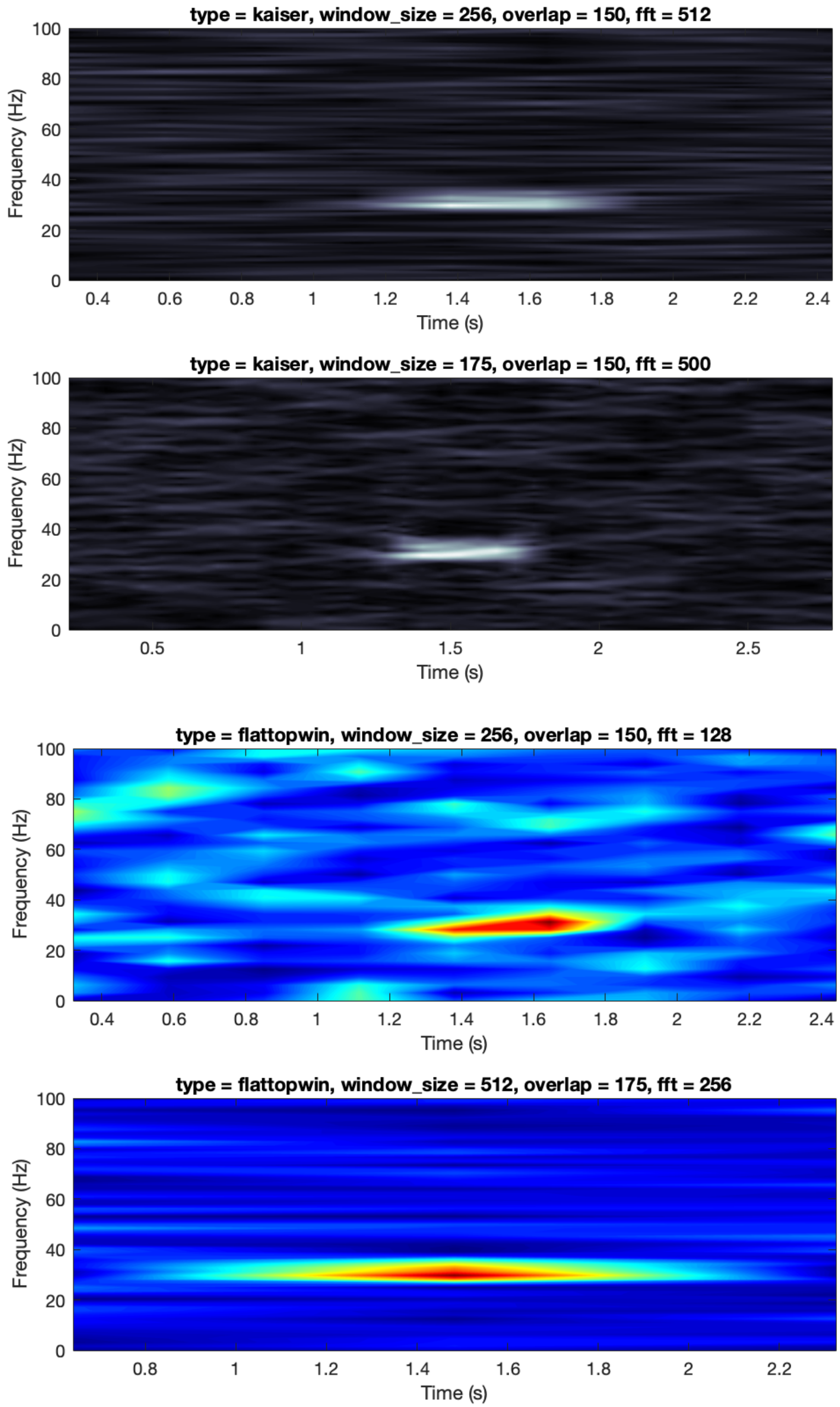


Figure 12: Test cases for kaiser and flattopwin.

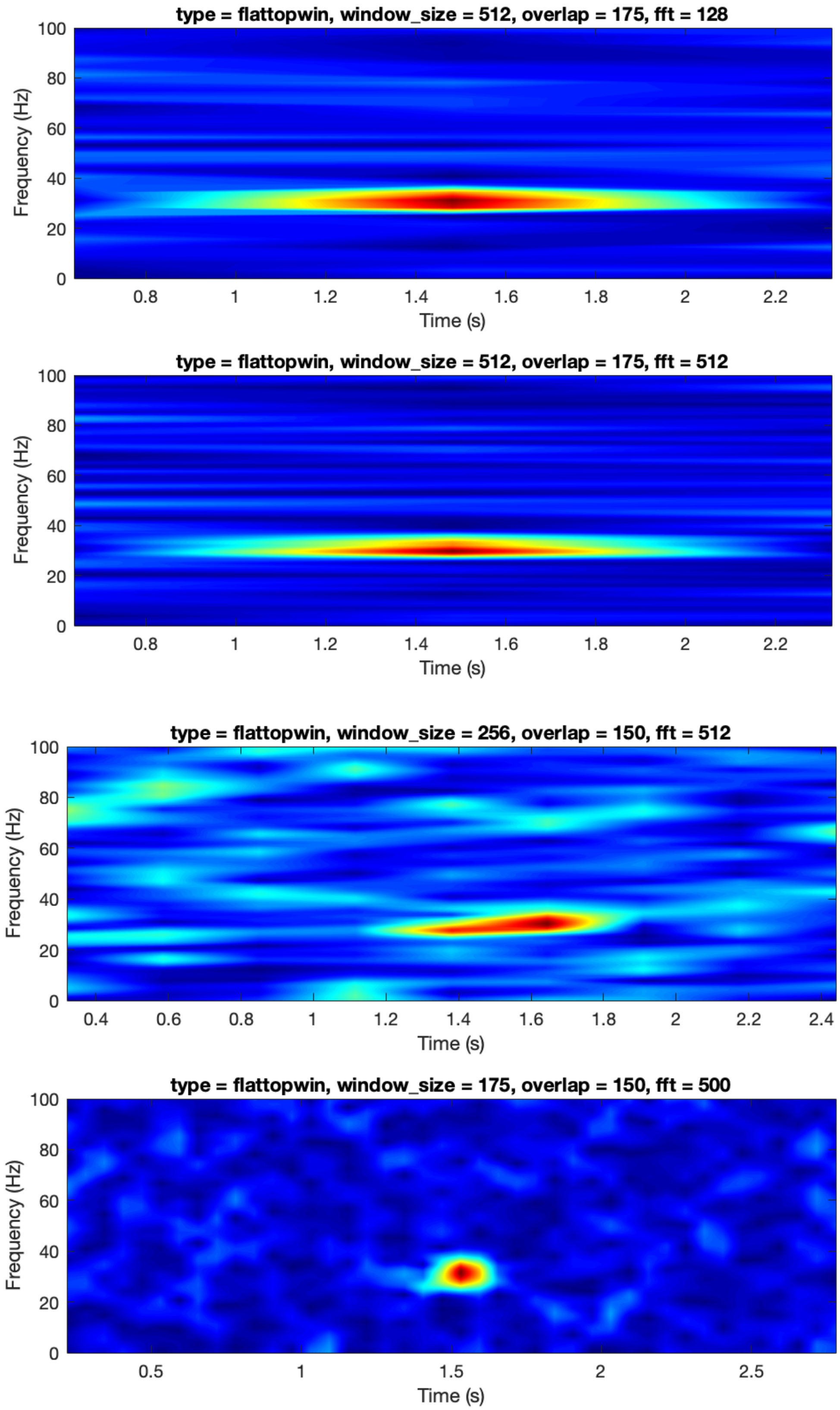


Figure 13: Test cases for flattopwin.

Clearly,
parameters: window_size = 175, overlap = 150 and fft_size = 500 (last plot for every window type) gives the best results for analyzing in every window type.

window_type: flattopwin and hann are good enough to find the location of transient activities. However, kaiser has only average accuracy. I also utilized different colormaps to increase readability. My preference is mostly jet colormap, because it provides high resolution activity. Overlap is an important metric because it restricts our other parameters, and window's overlapping tolerance.

The higher window_size, the lower accuracy for finding transient activity.

The higher fft_size, the lower frequency band.

RESEARCH

I did my research, but couldn't find any useful information or code to employ in my current study except Mike Cohen's study materials. Therefore, few of needed cites are provided in the relevant pages.

CODE

Code files are shared with the report path as a .m file.

COMMENTS ABOUT THE CODE

I have no issue with coding due to my computer engineering coursework in the university. Therefore, the slightly harder part was explaining the results.

RESULTS & COMMENTS ABOUT THE RESULTS

As can be seen from many explanations I made in the questions, I mostly dived directly into the problem and made predictions that were already written in the question, and I got results that were in line with what was expected in the question without being aware of what was said in the question as an addition to the main task such as existence of noises. I believe the results I found are correct because they are all consistent with the results I expected.

In the end, I would rather solve a problem with real brain signals rather than work with artificial signals.