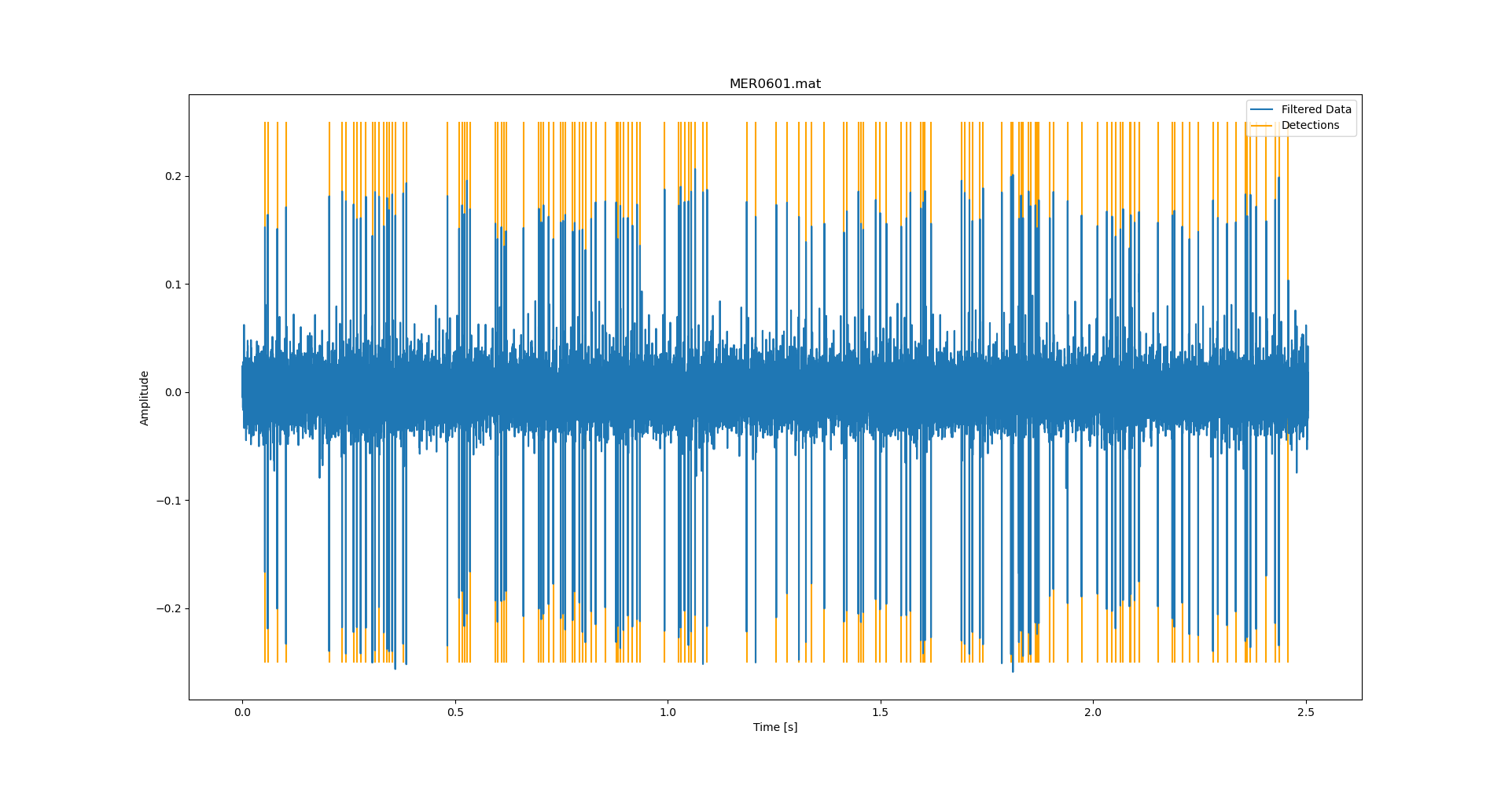
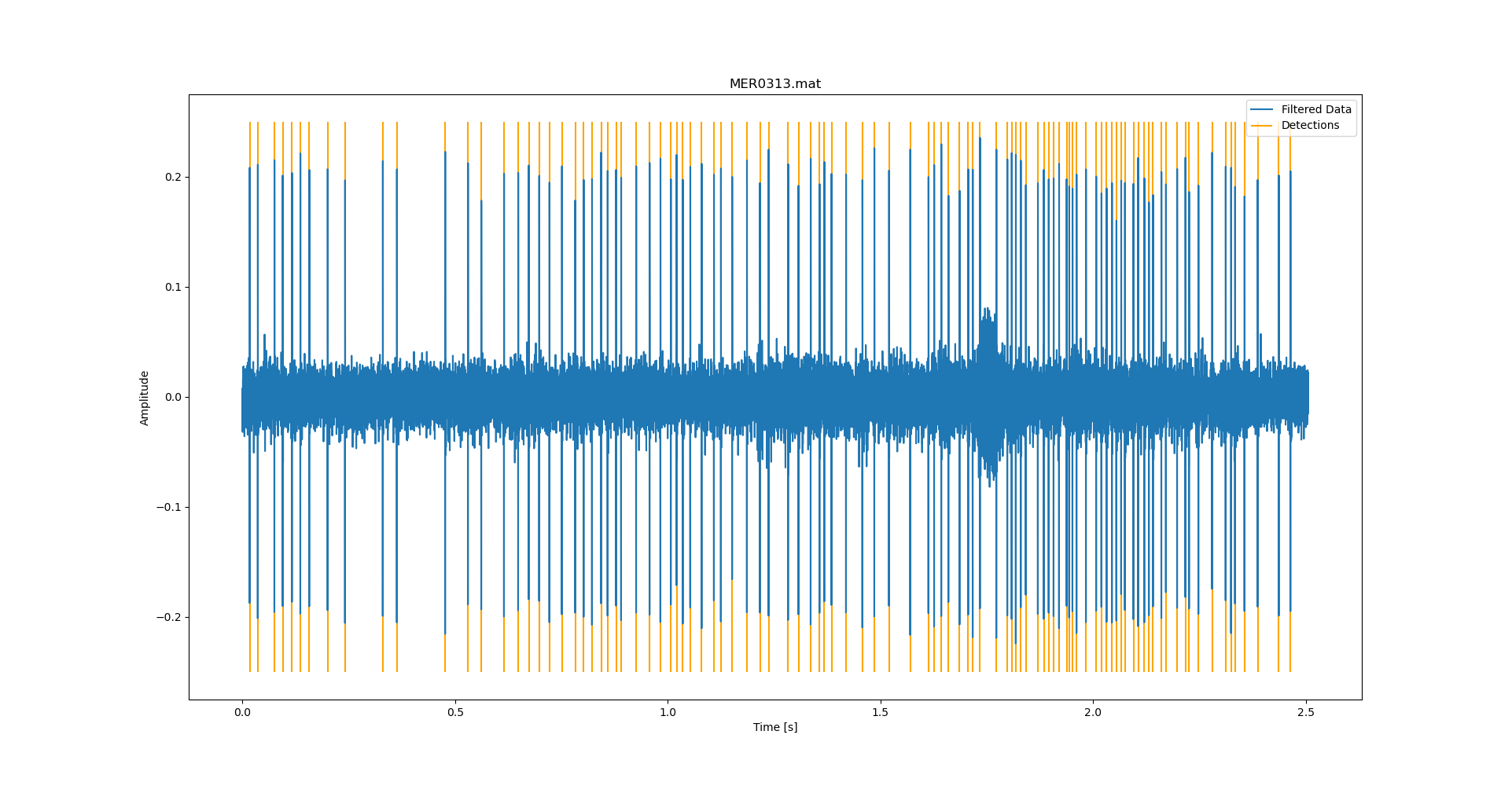
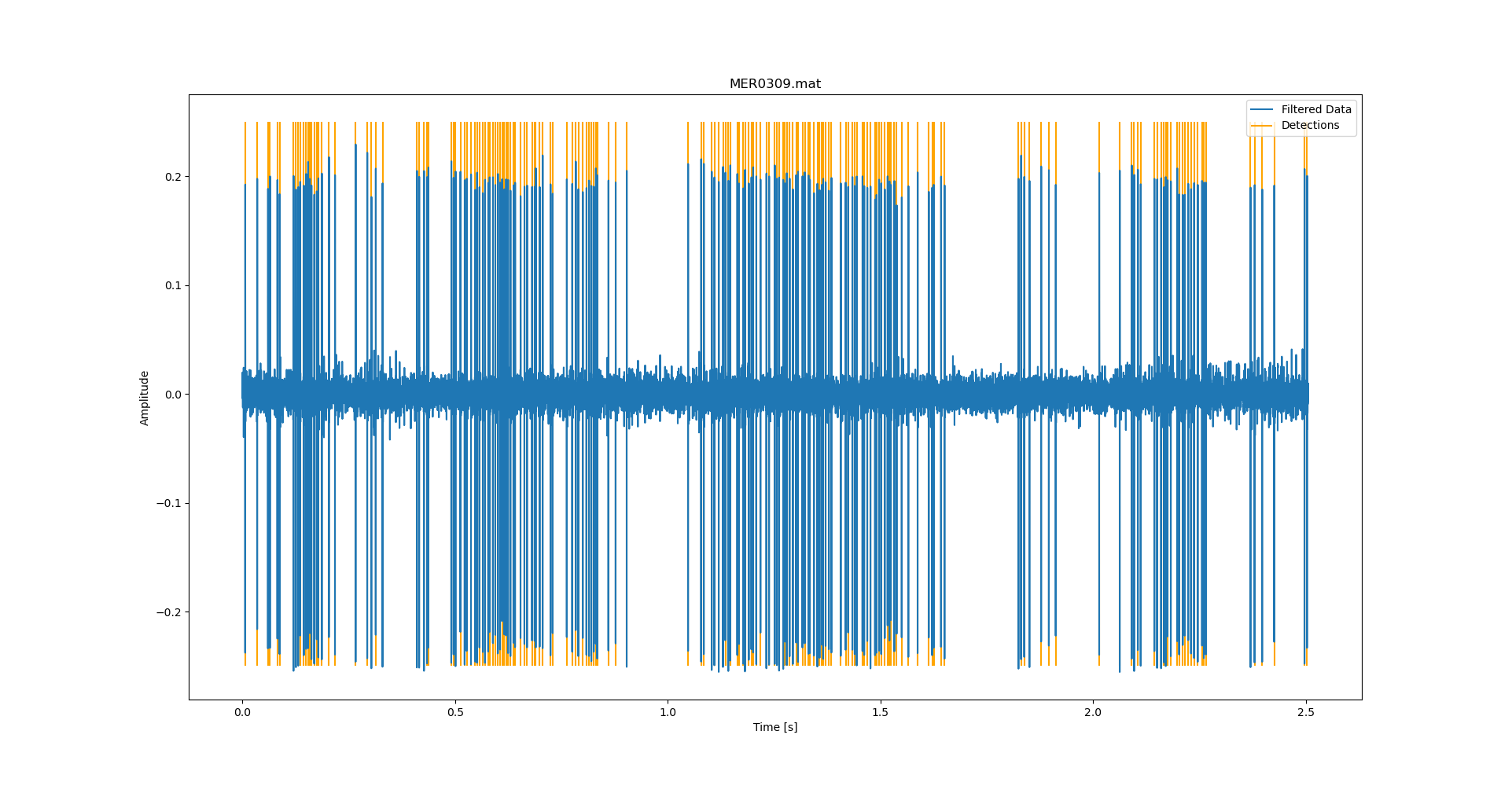
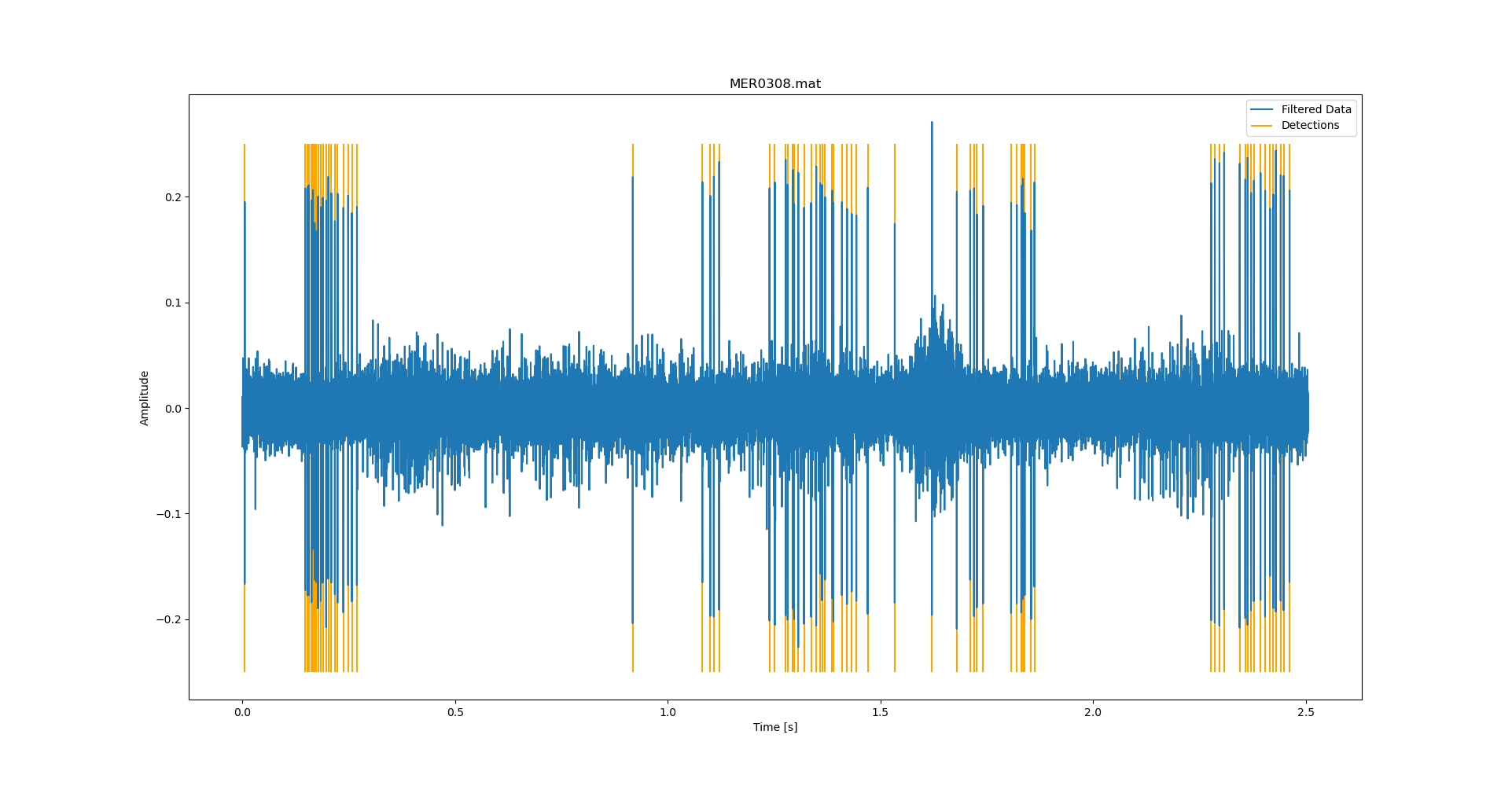
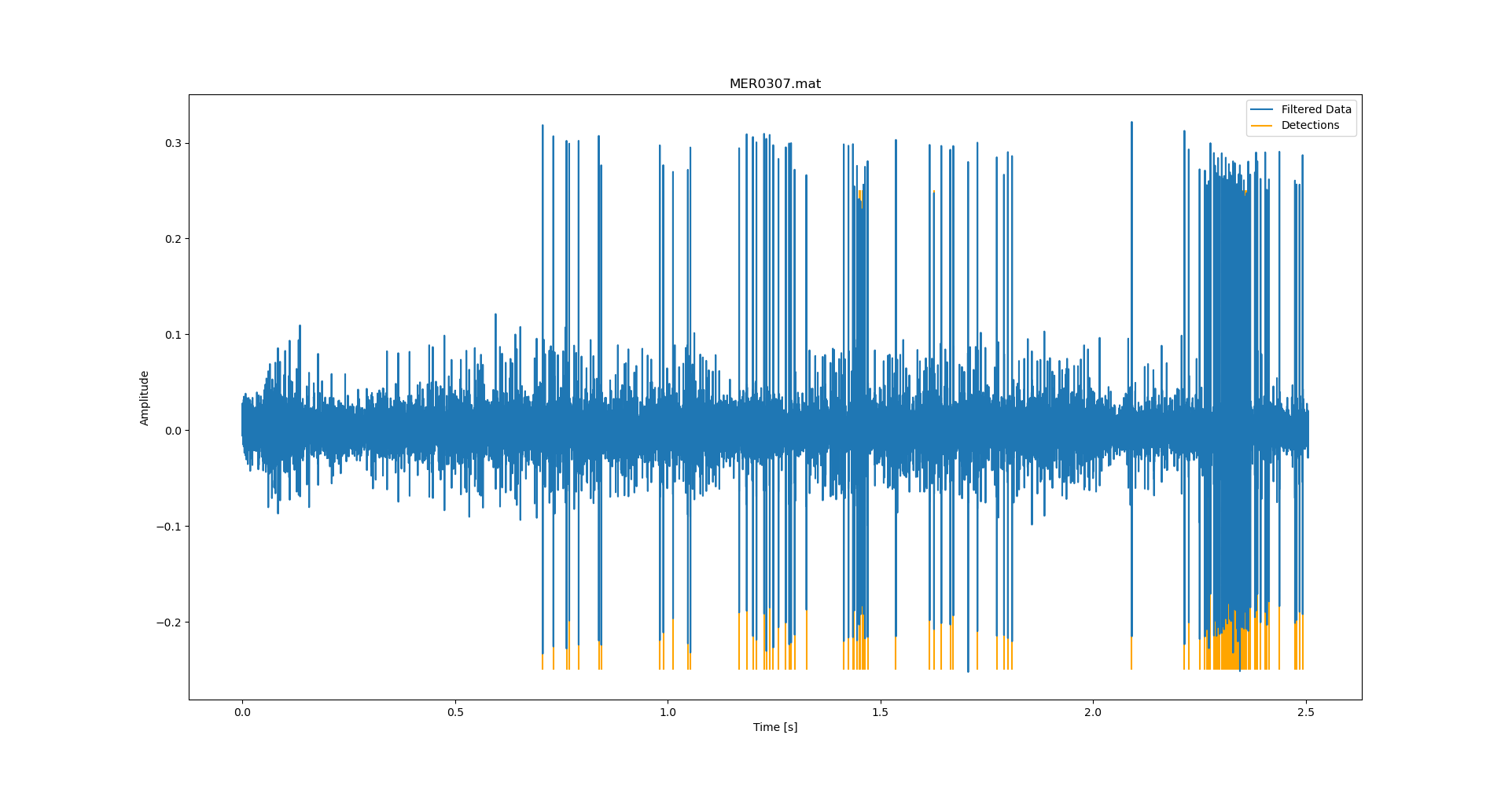
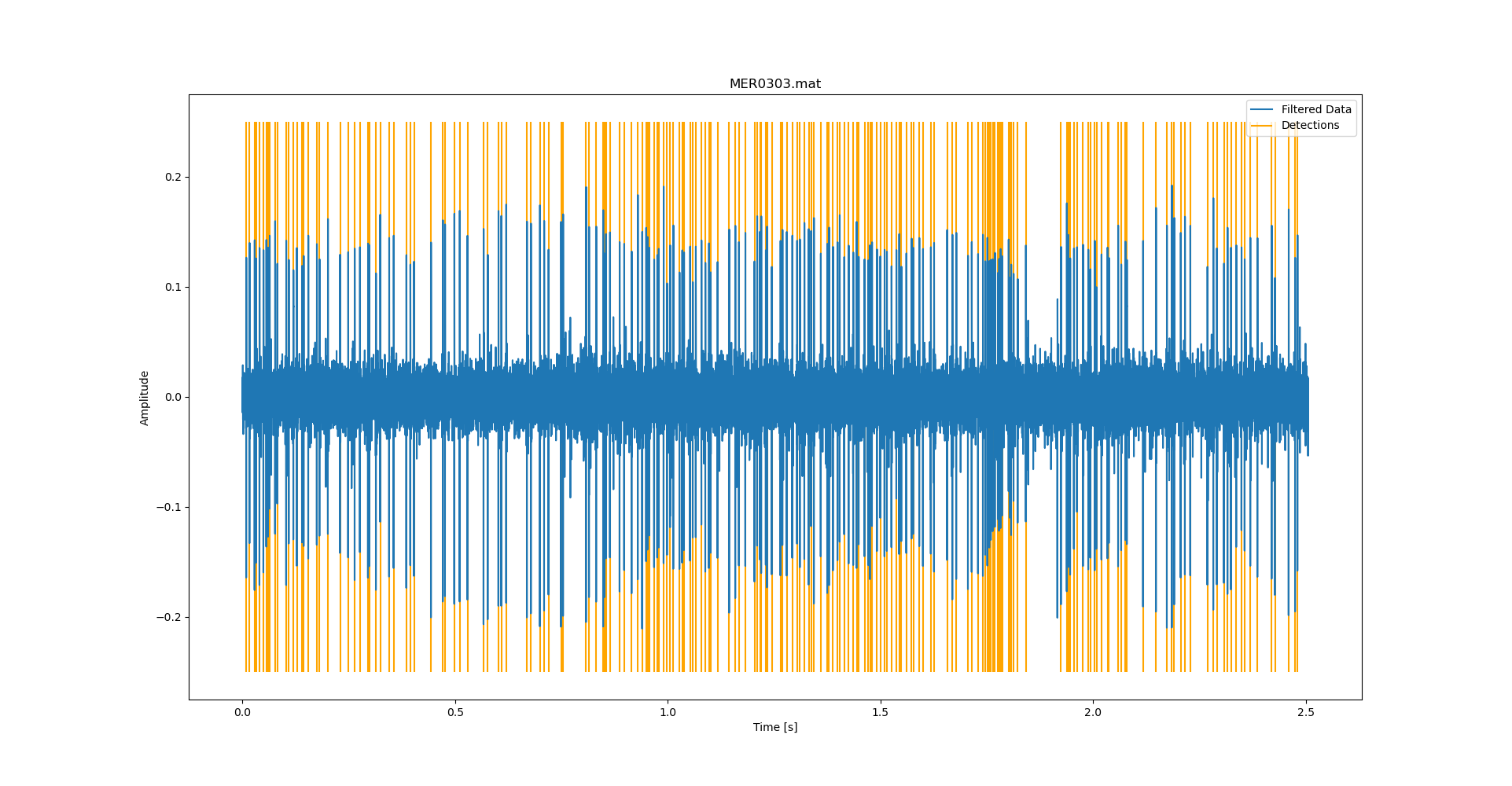
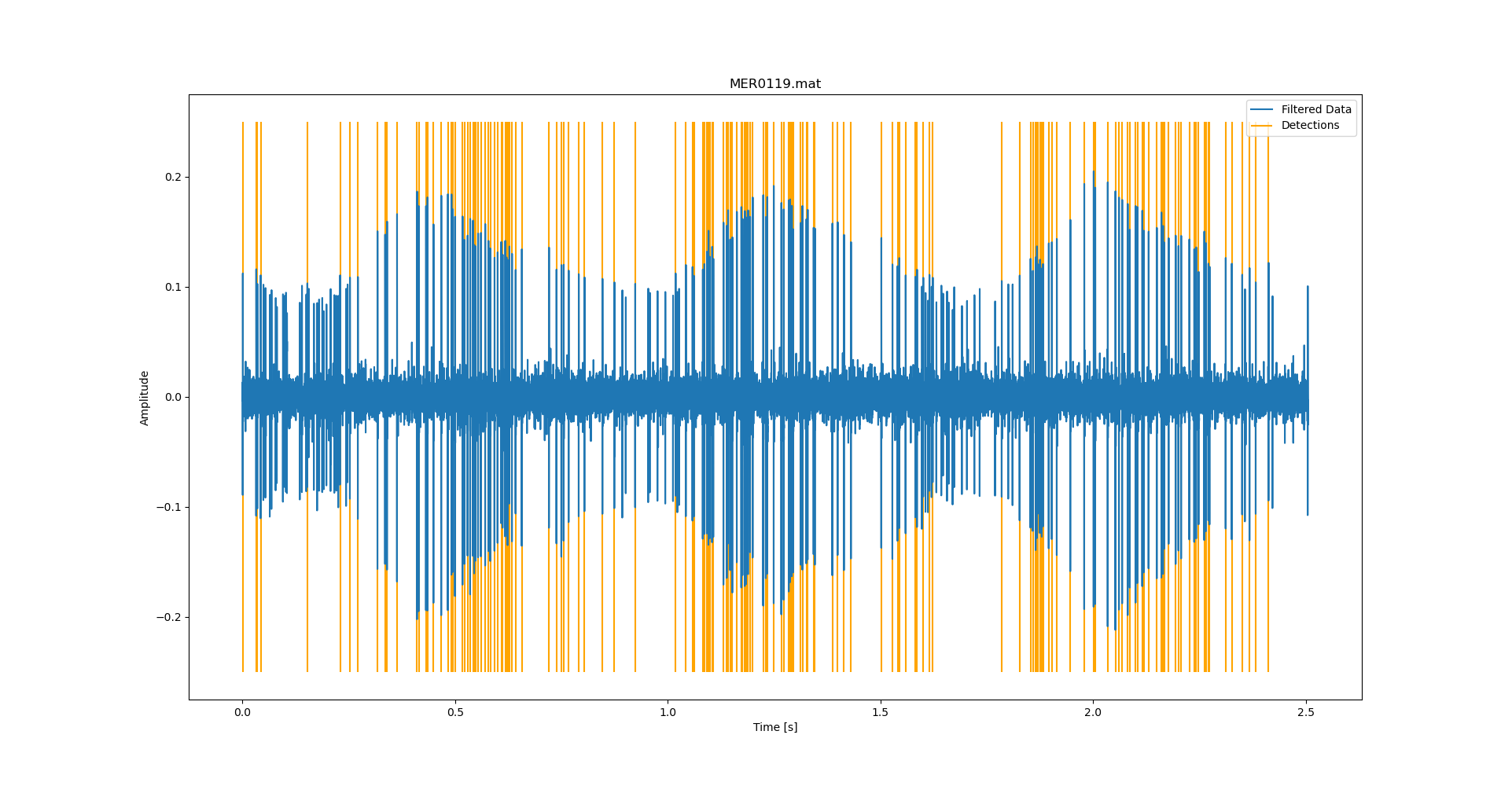
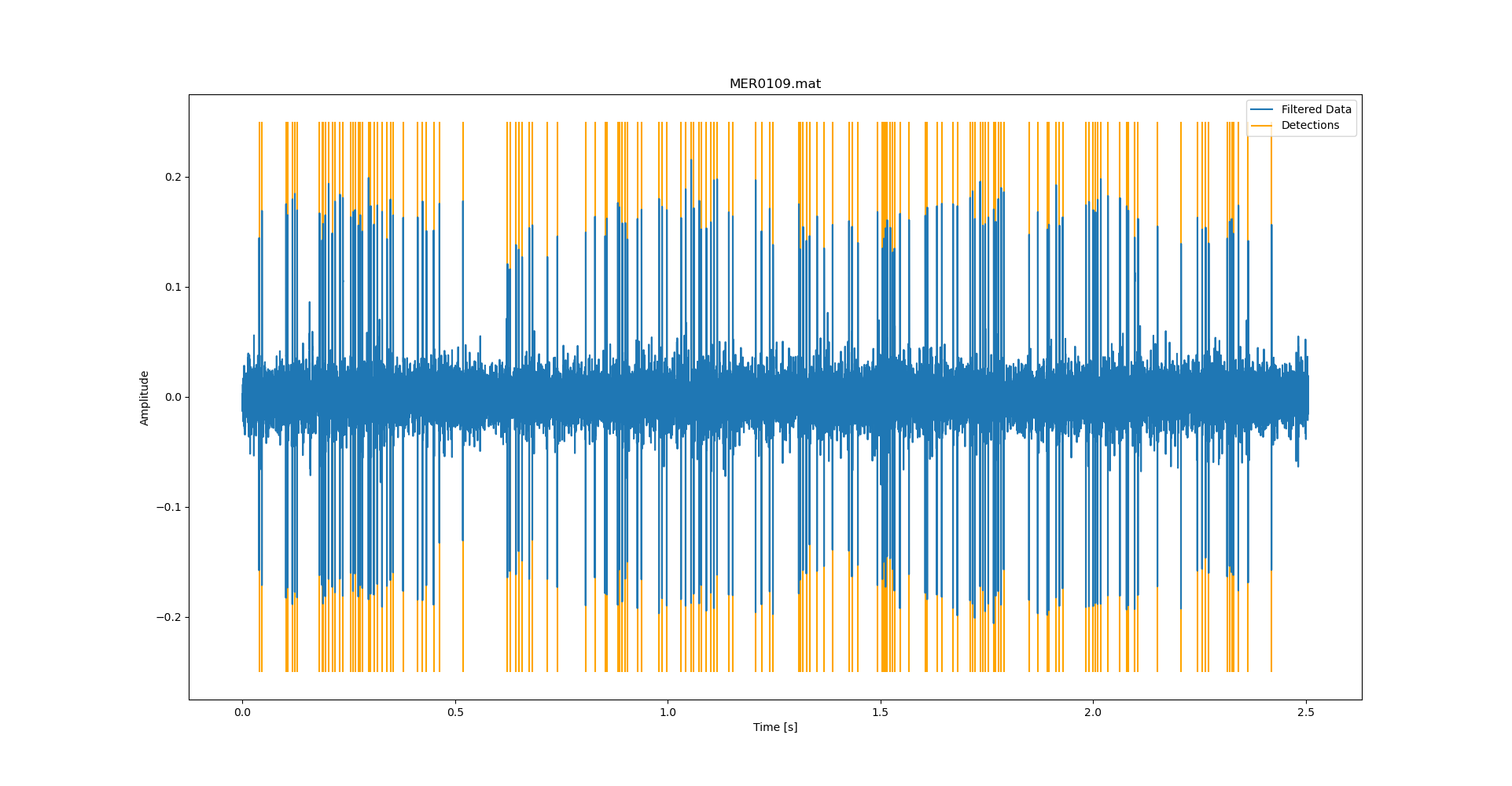
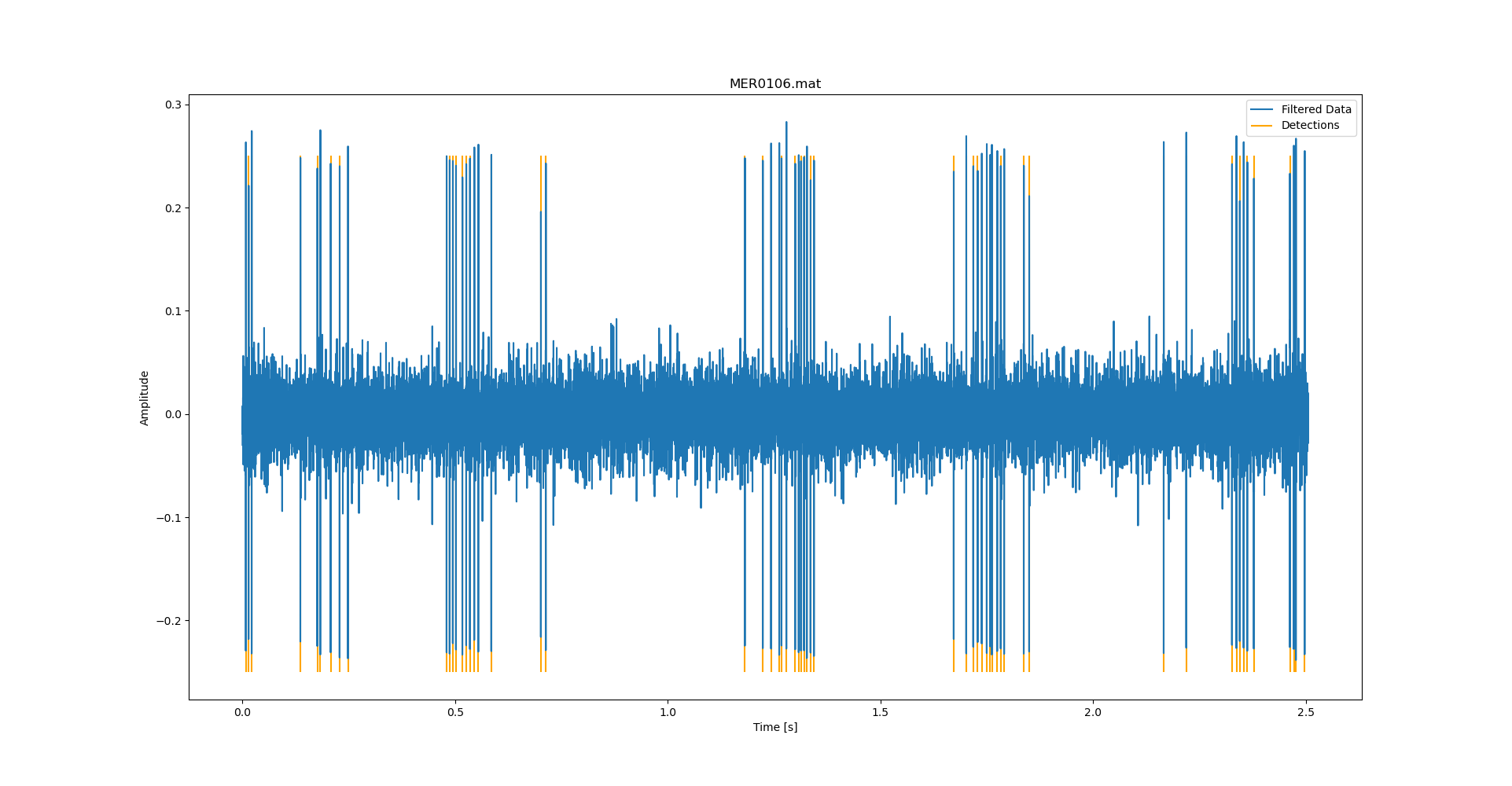
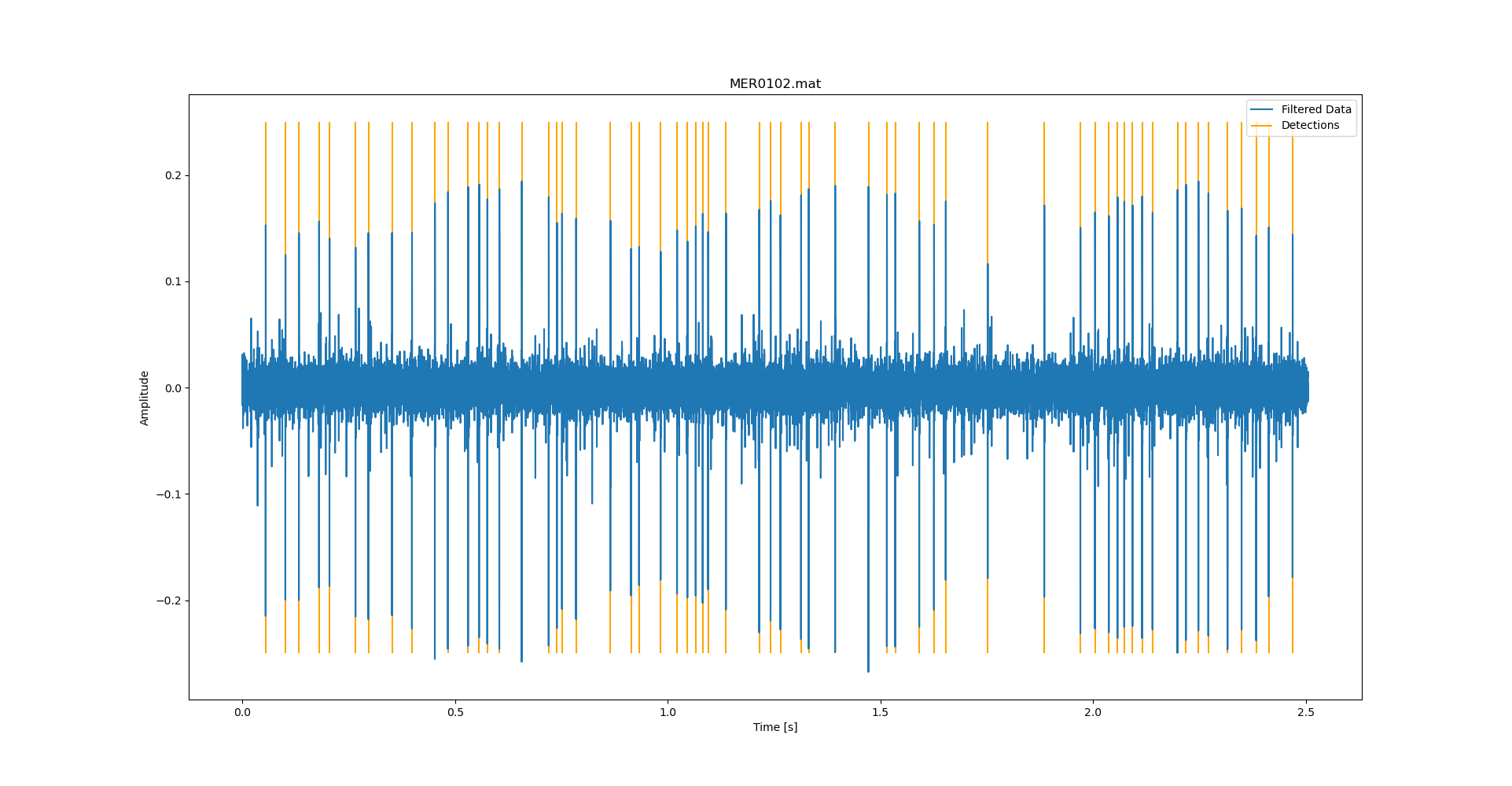
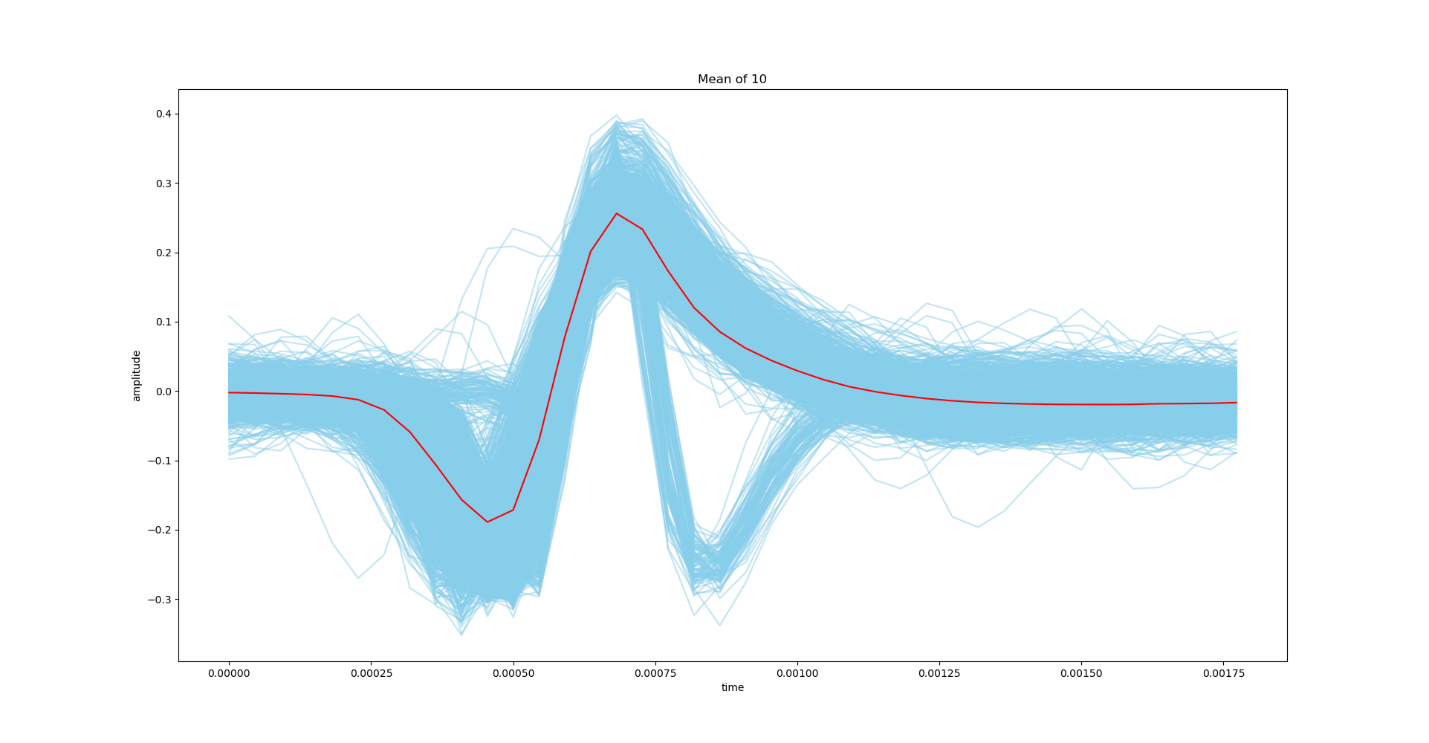
Mikhail Mayers

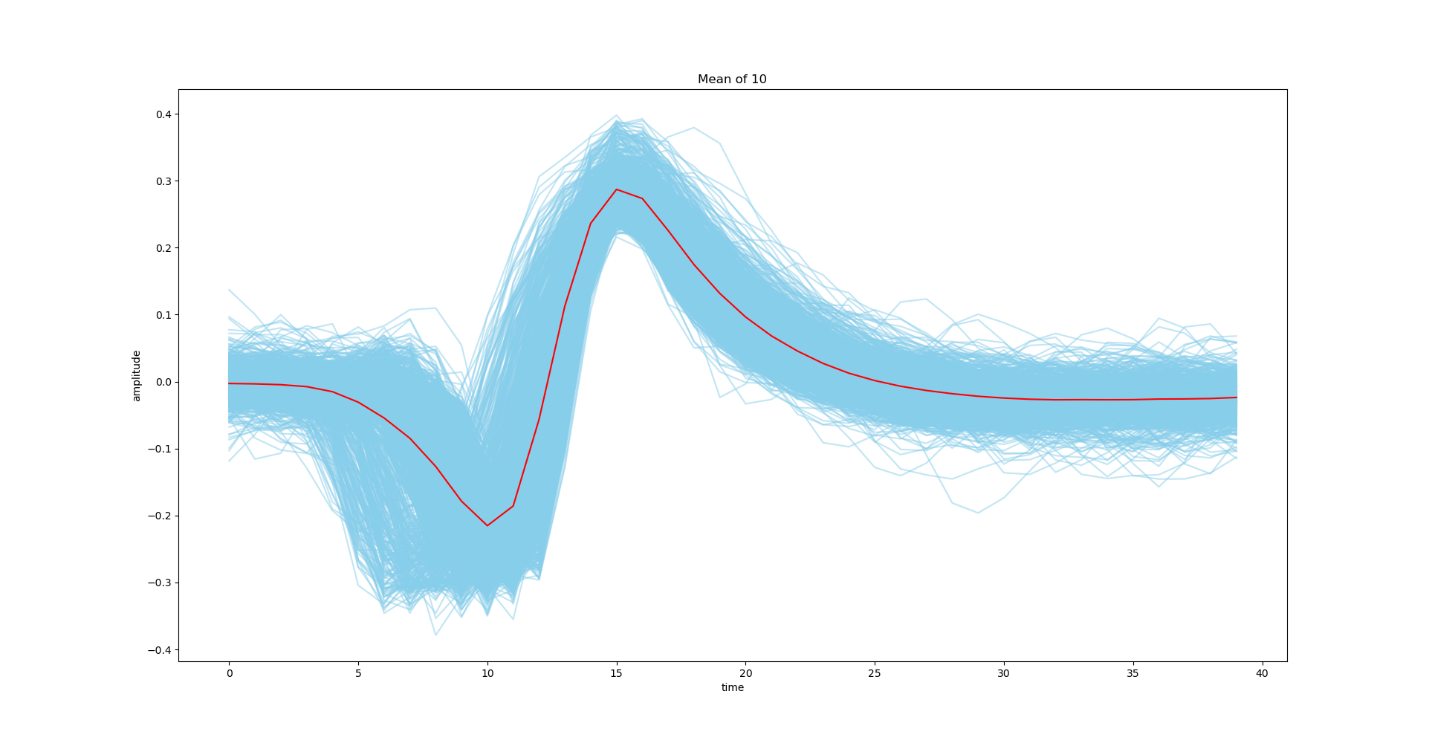
2/10/2021

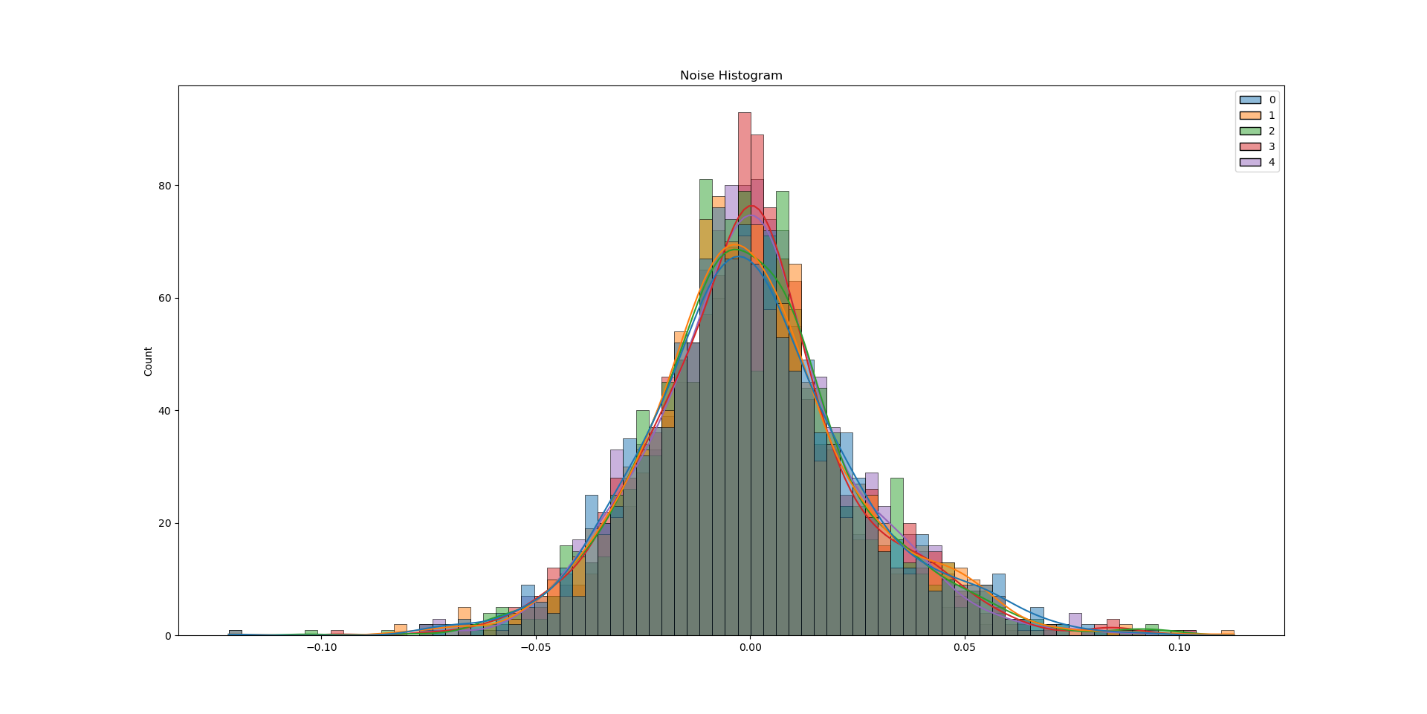
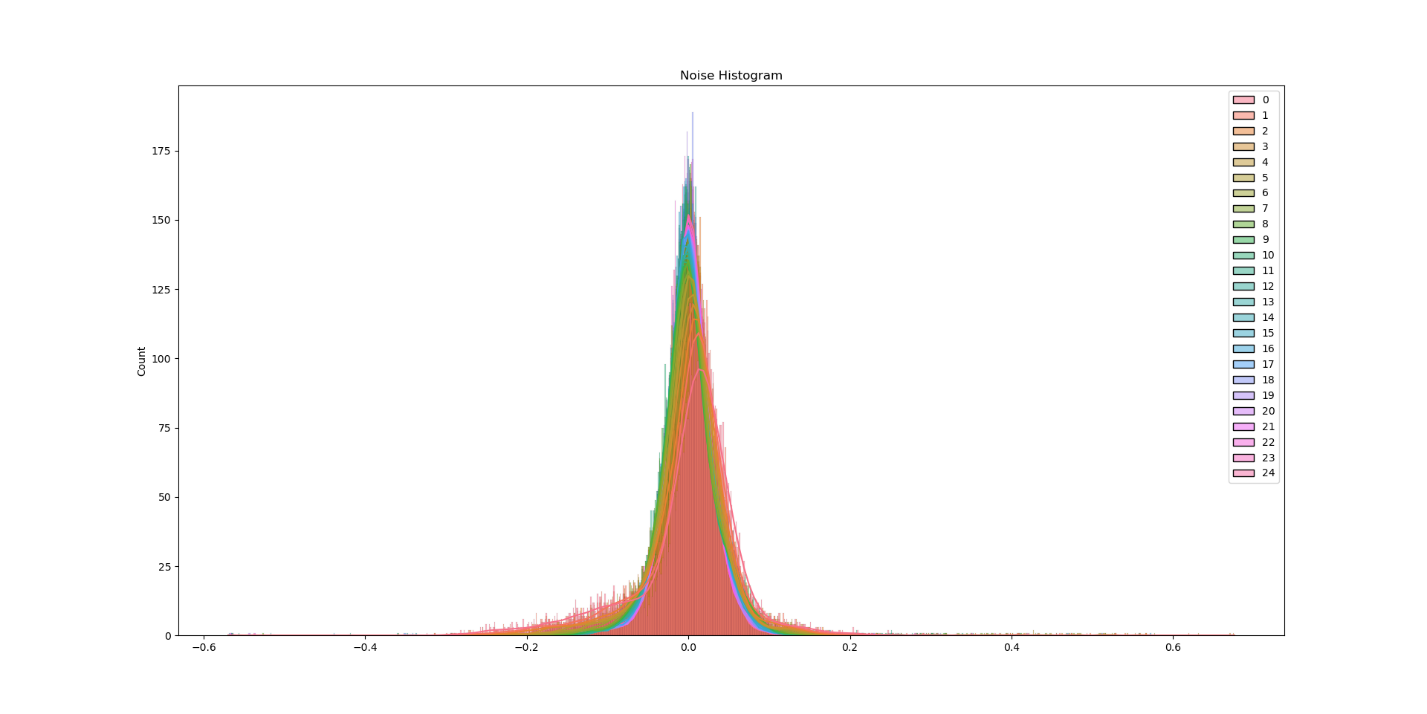
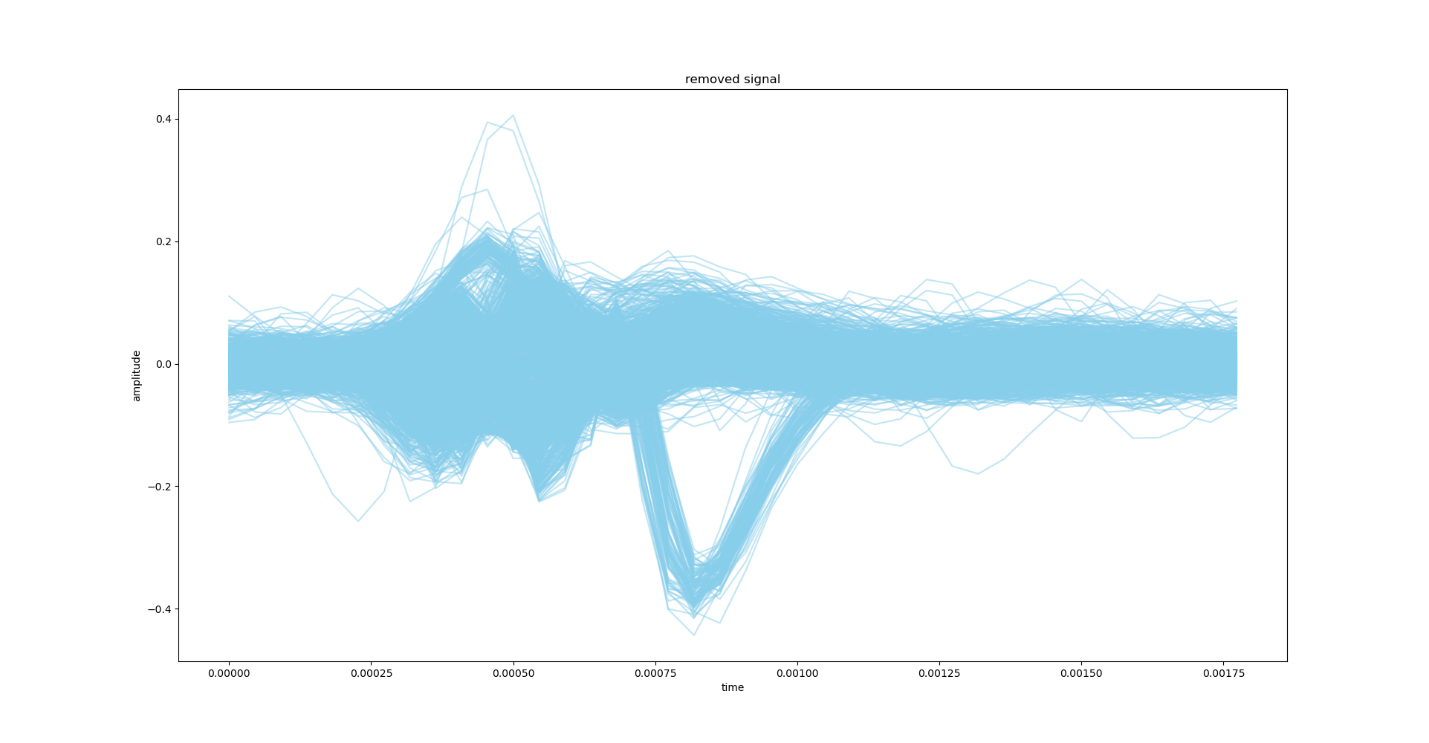
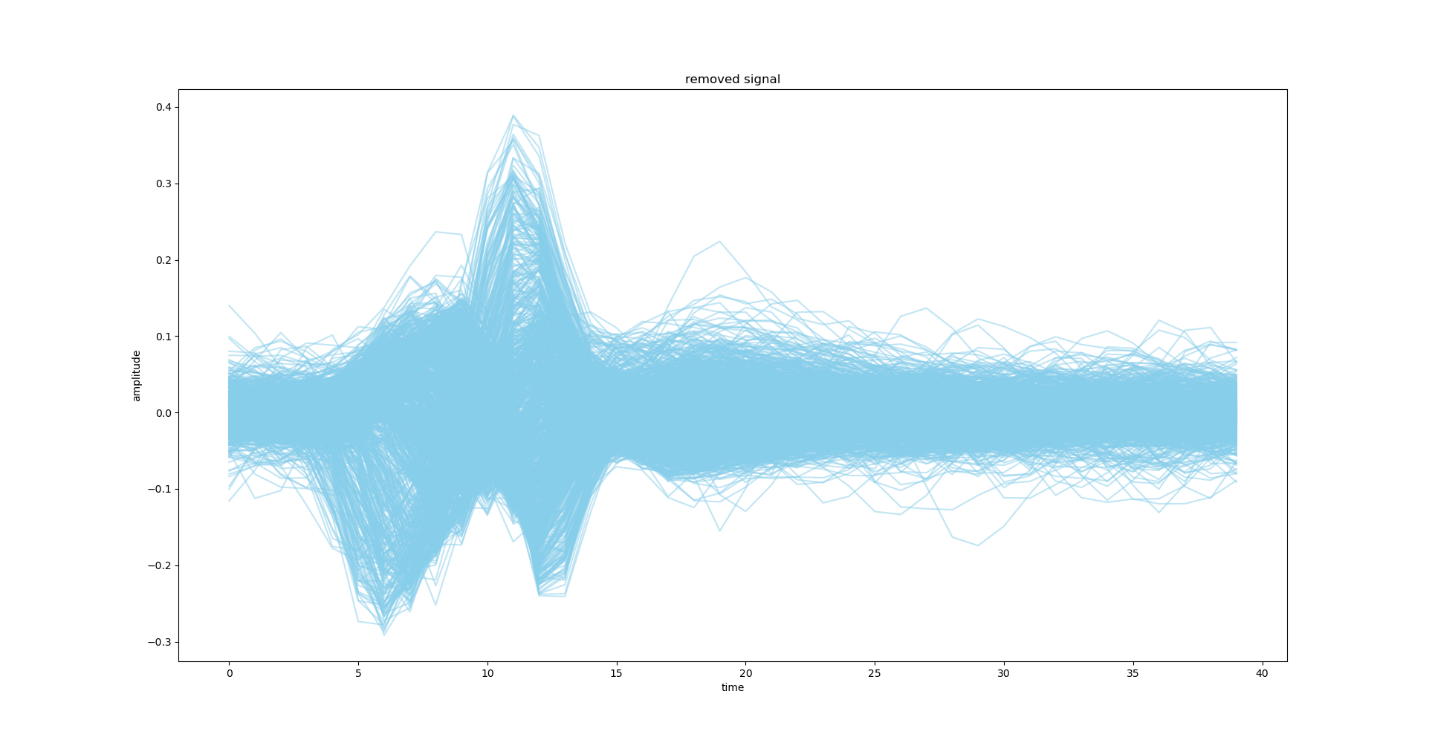
hw4

1. The initial heuristic approach I created was just my HW2 assignment from last term. We had the same assignment last term, with one waveform called Henderson. Luckily, I write all my code modularly, so it ended up being plug and play. I ended up removing unnecessary plots to make it run faster, but I didn’t have to change much more than that.

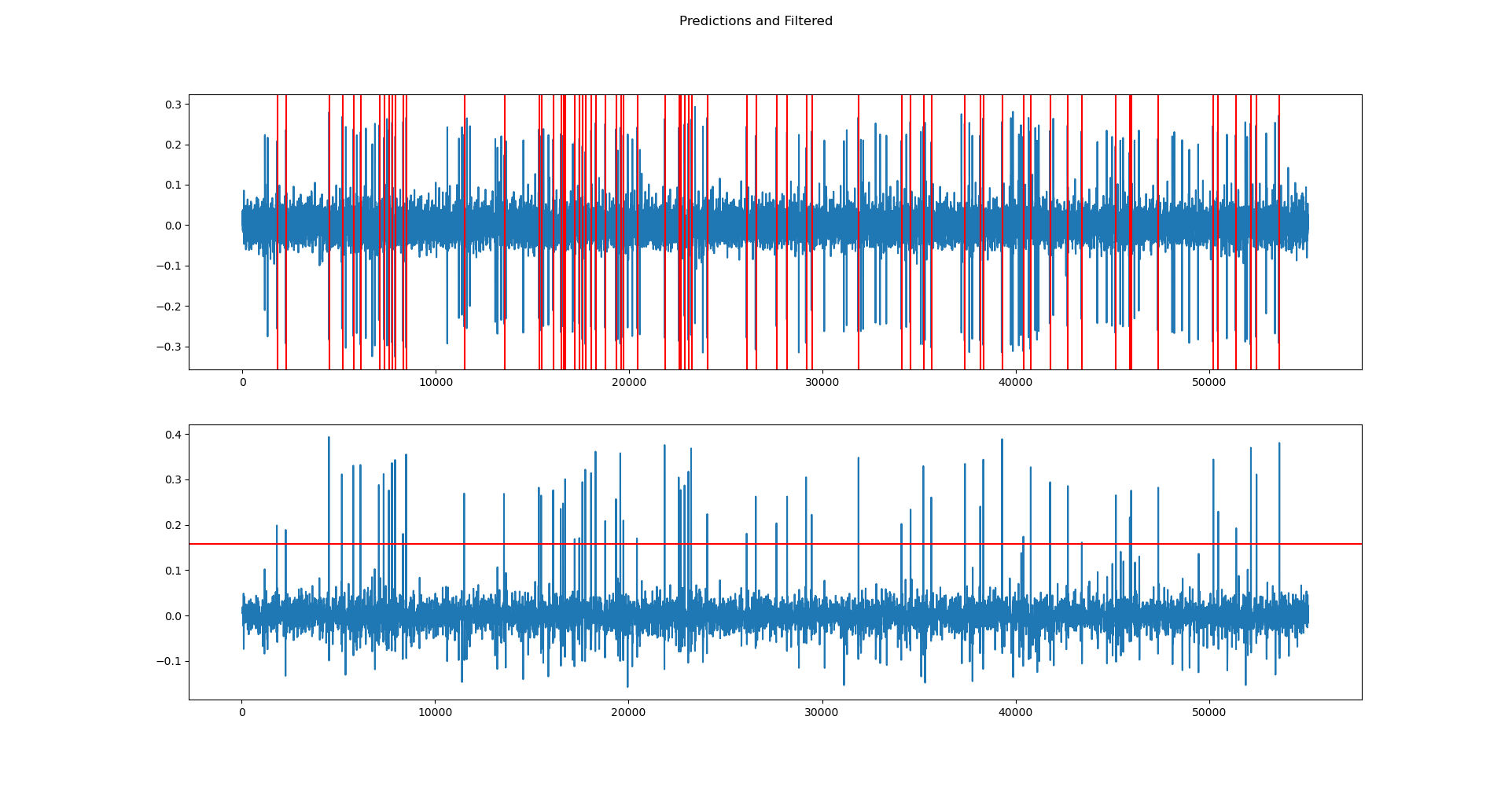
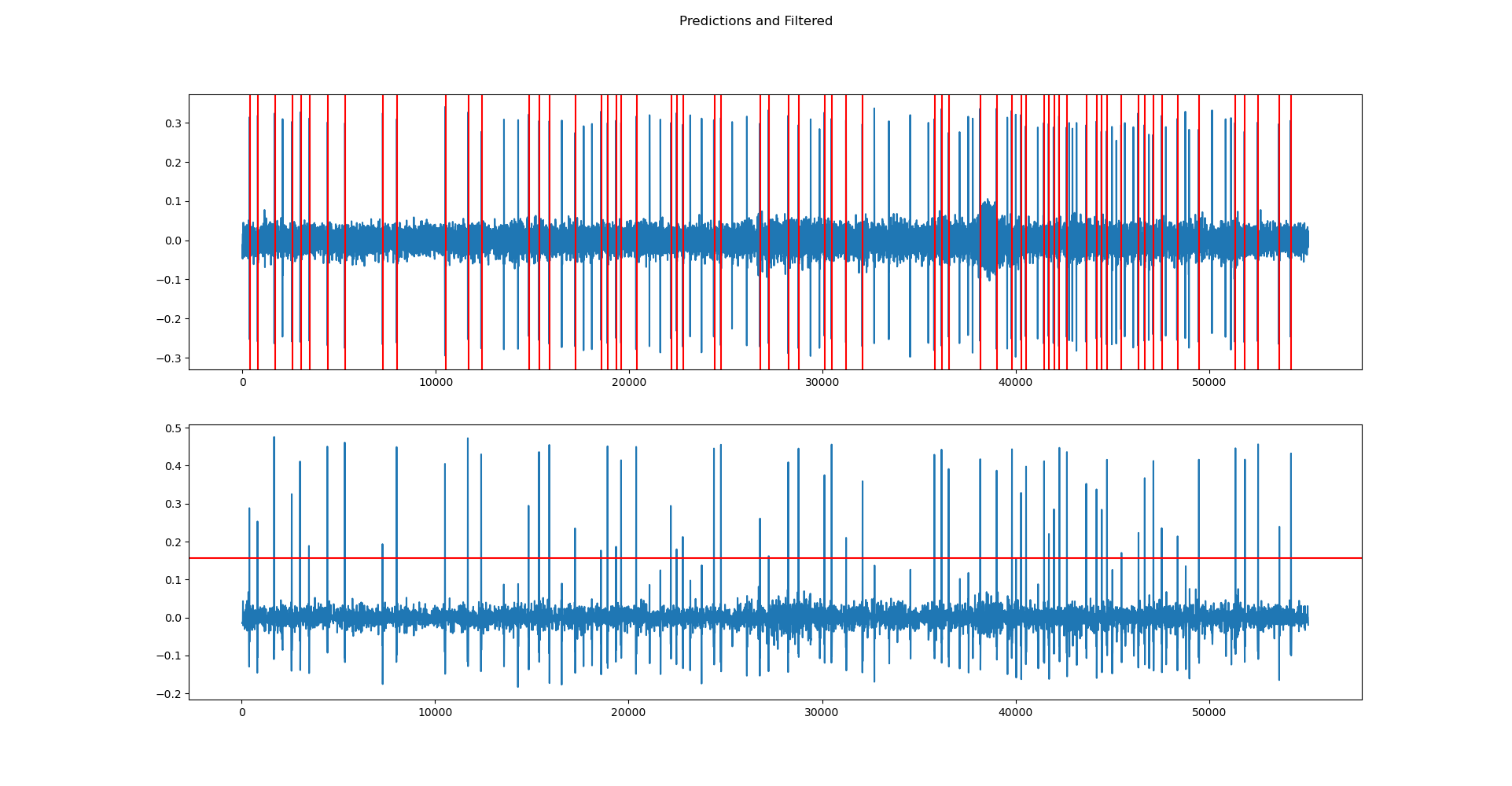
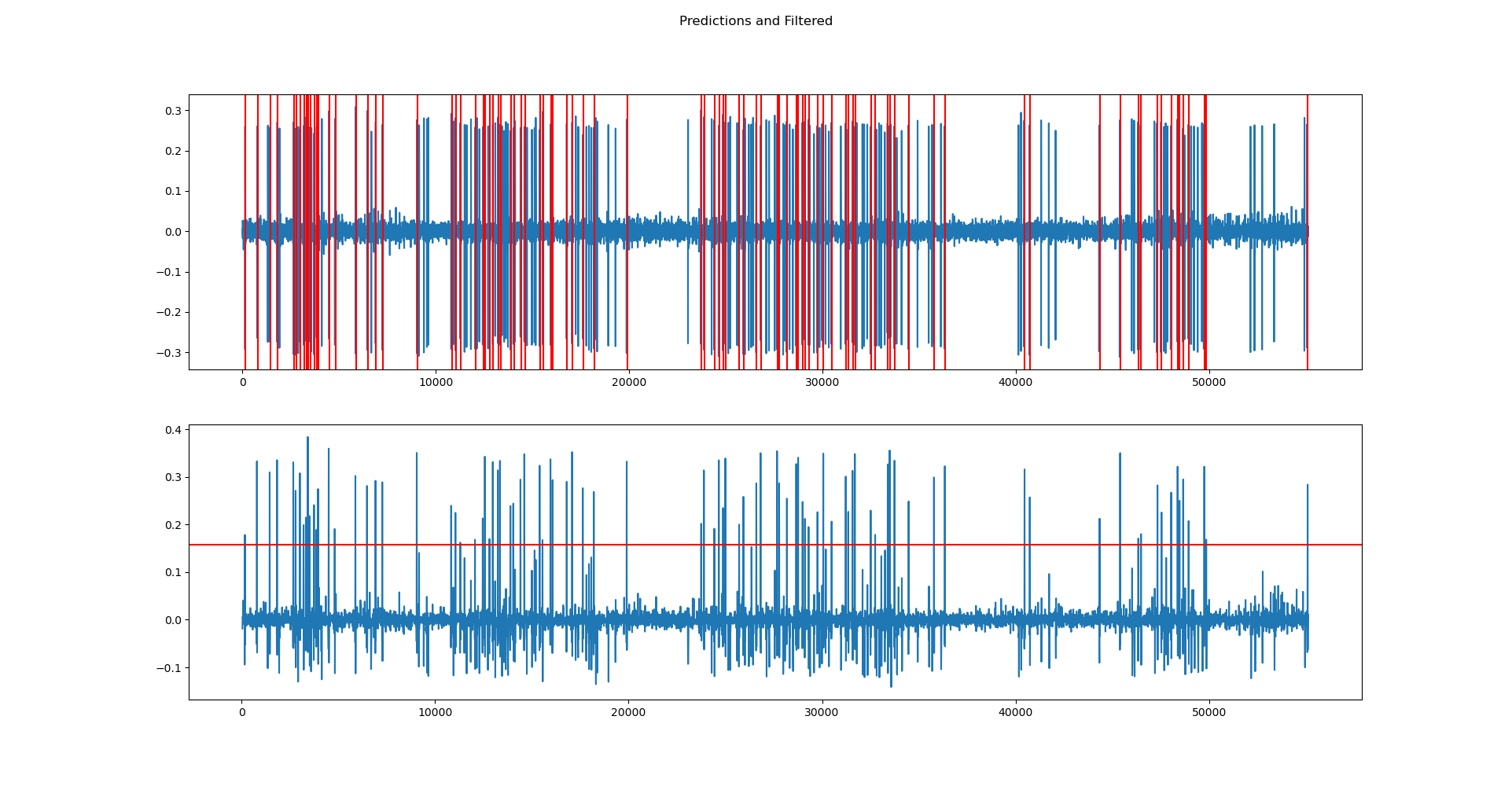
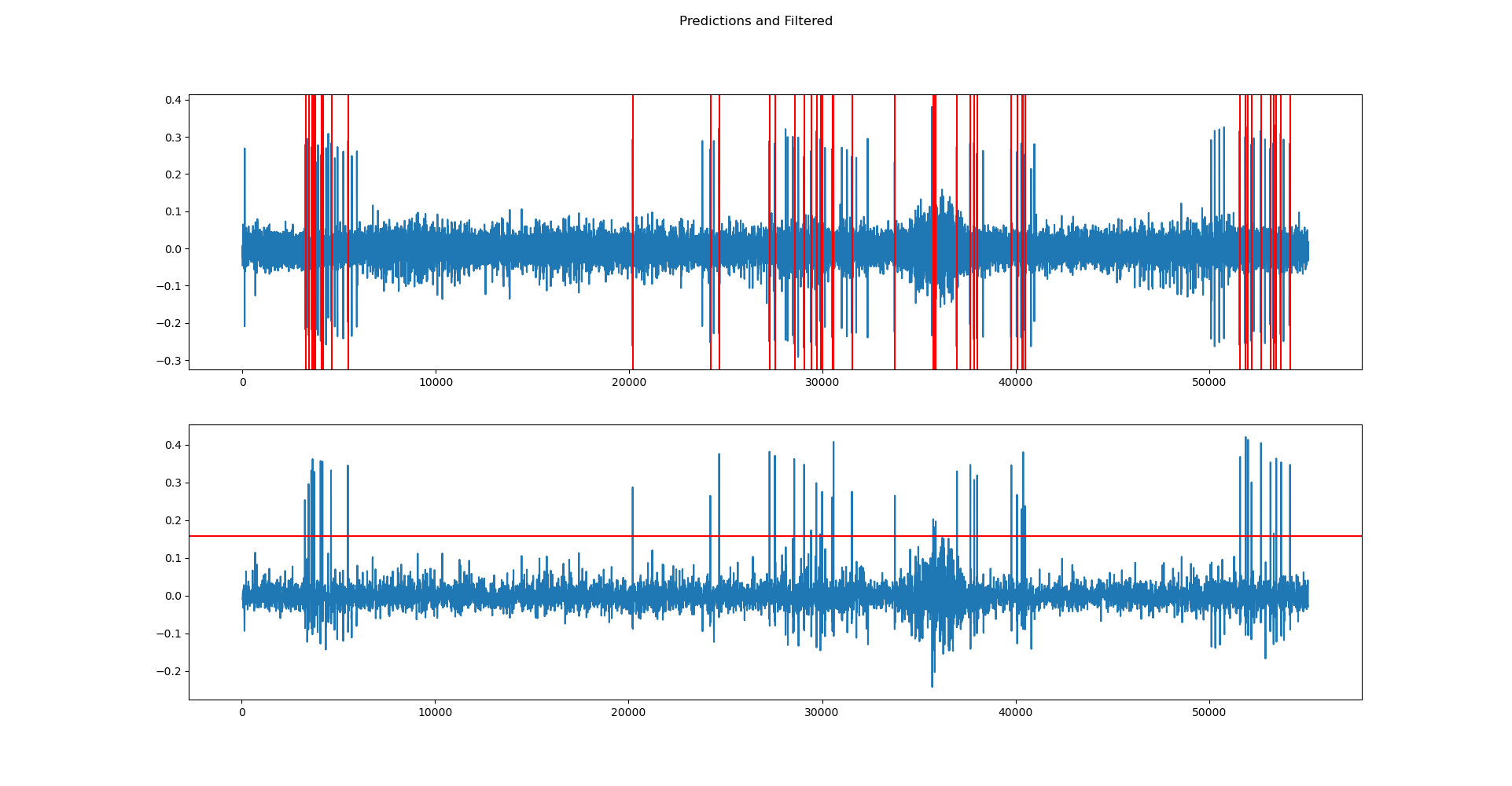
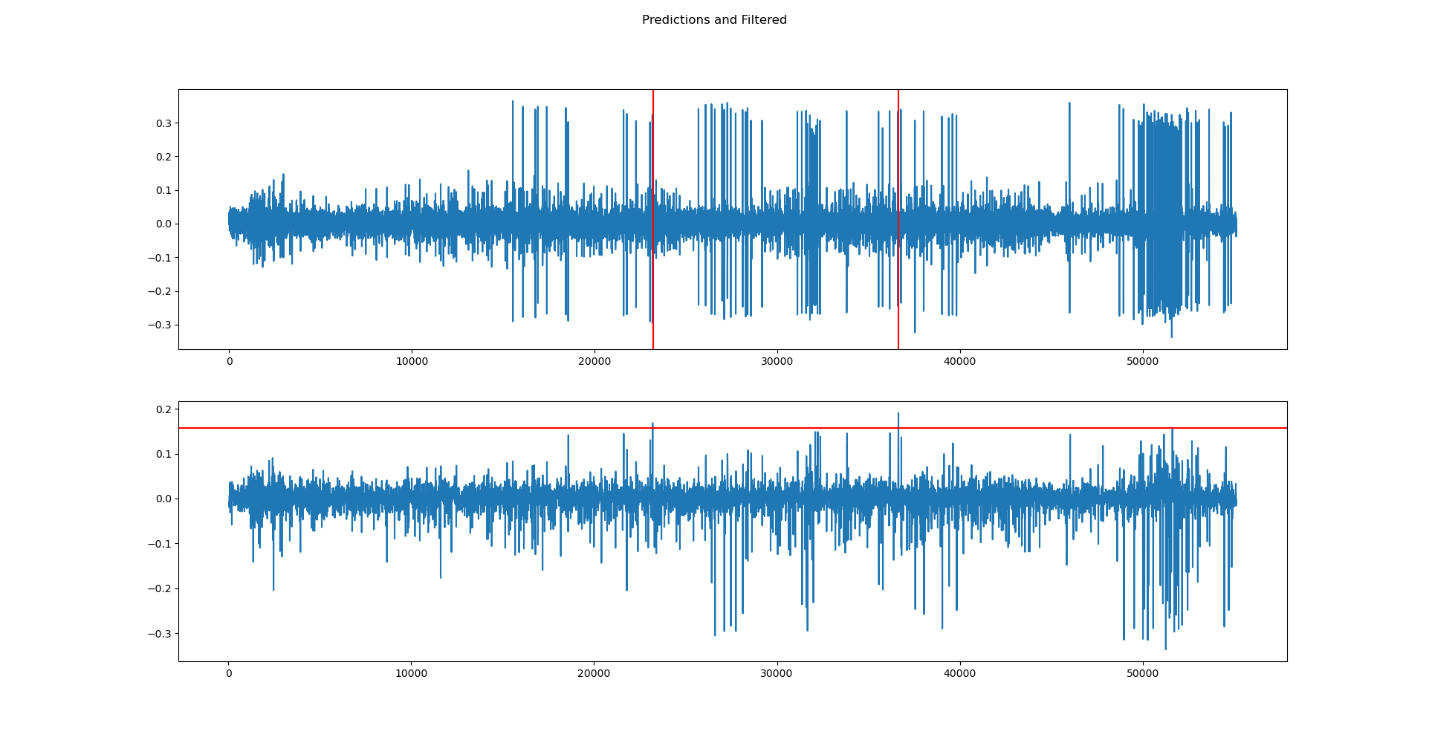
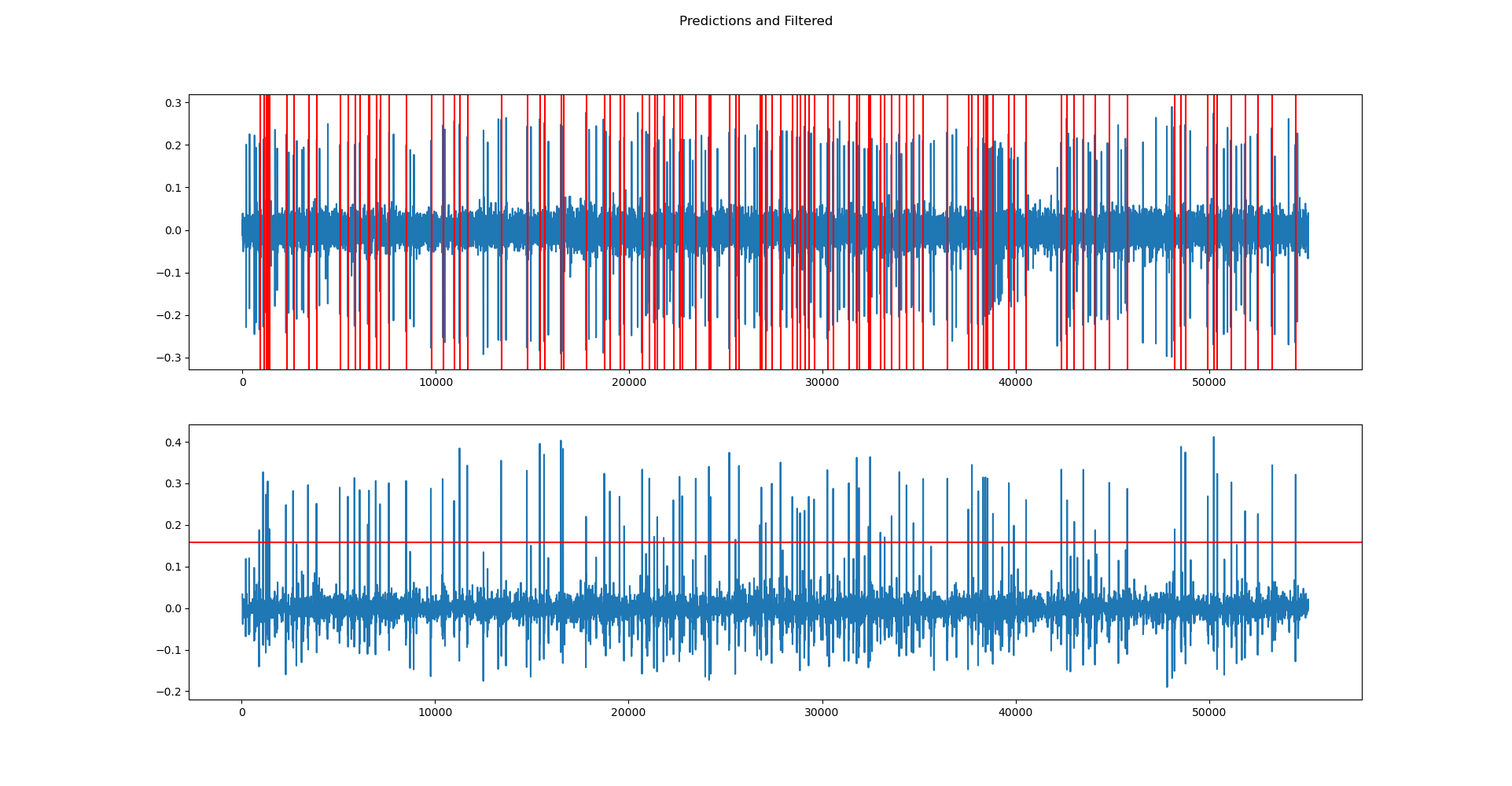
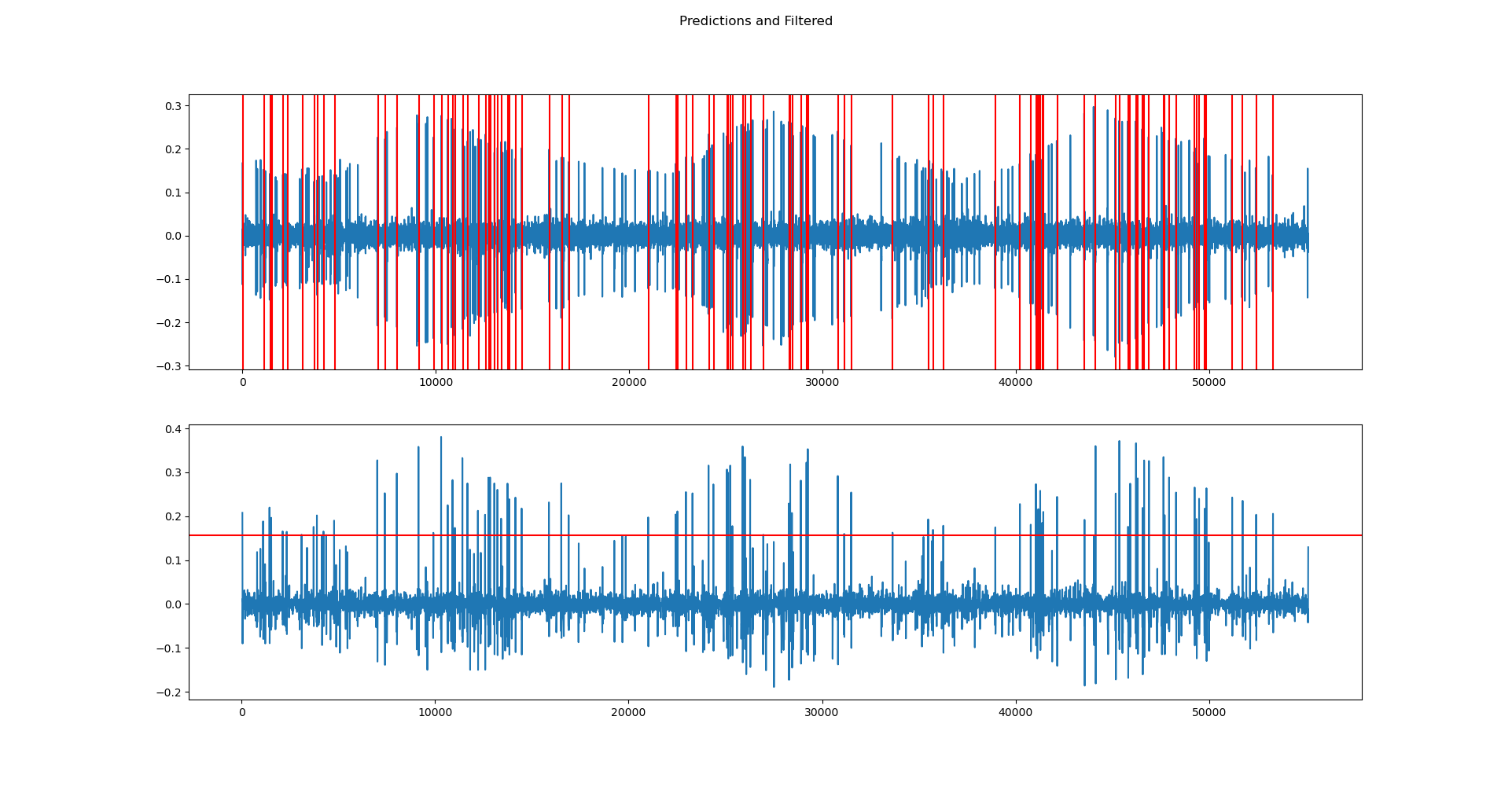
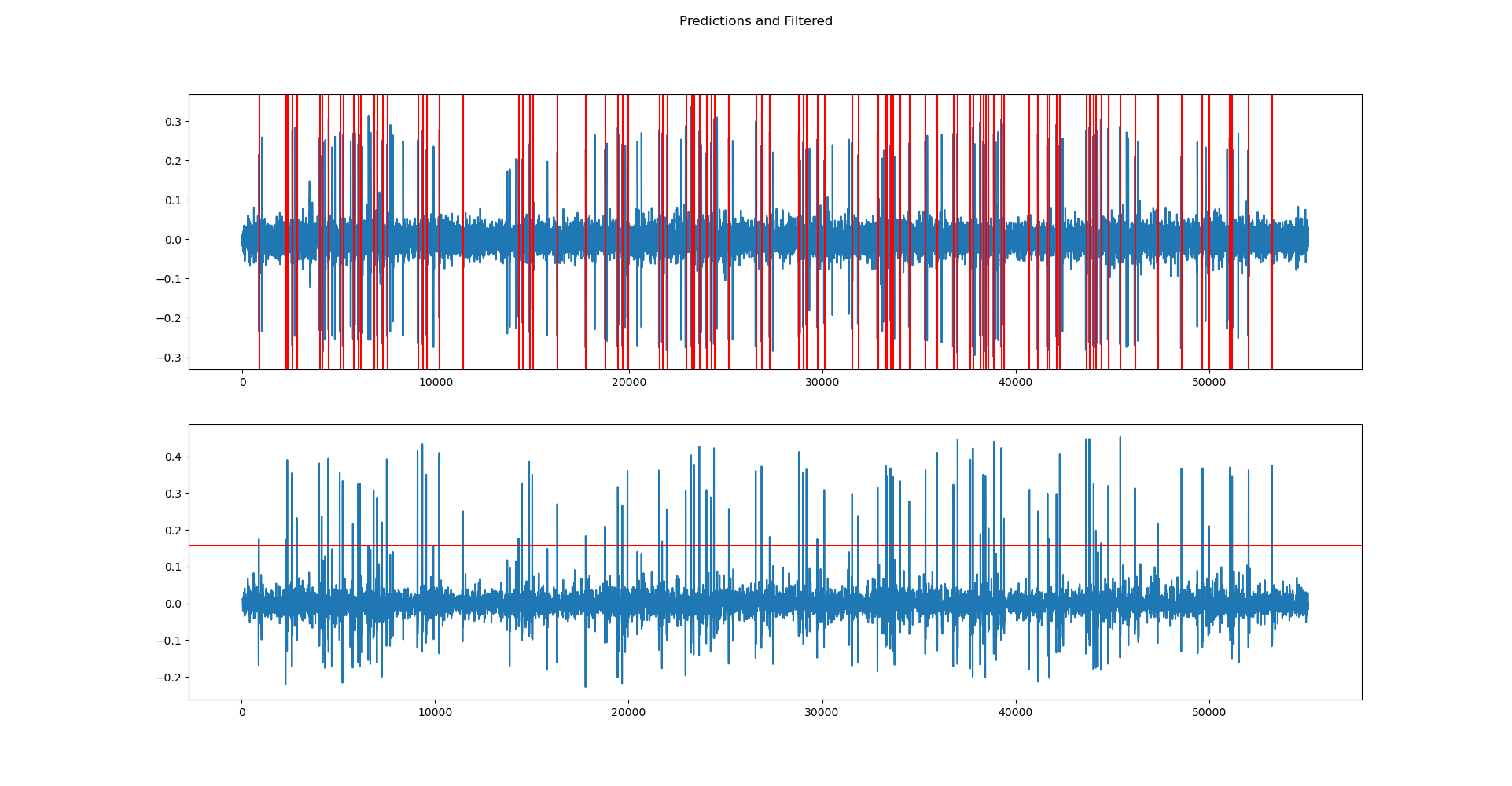
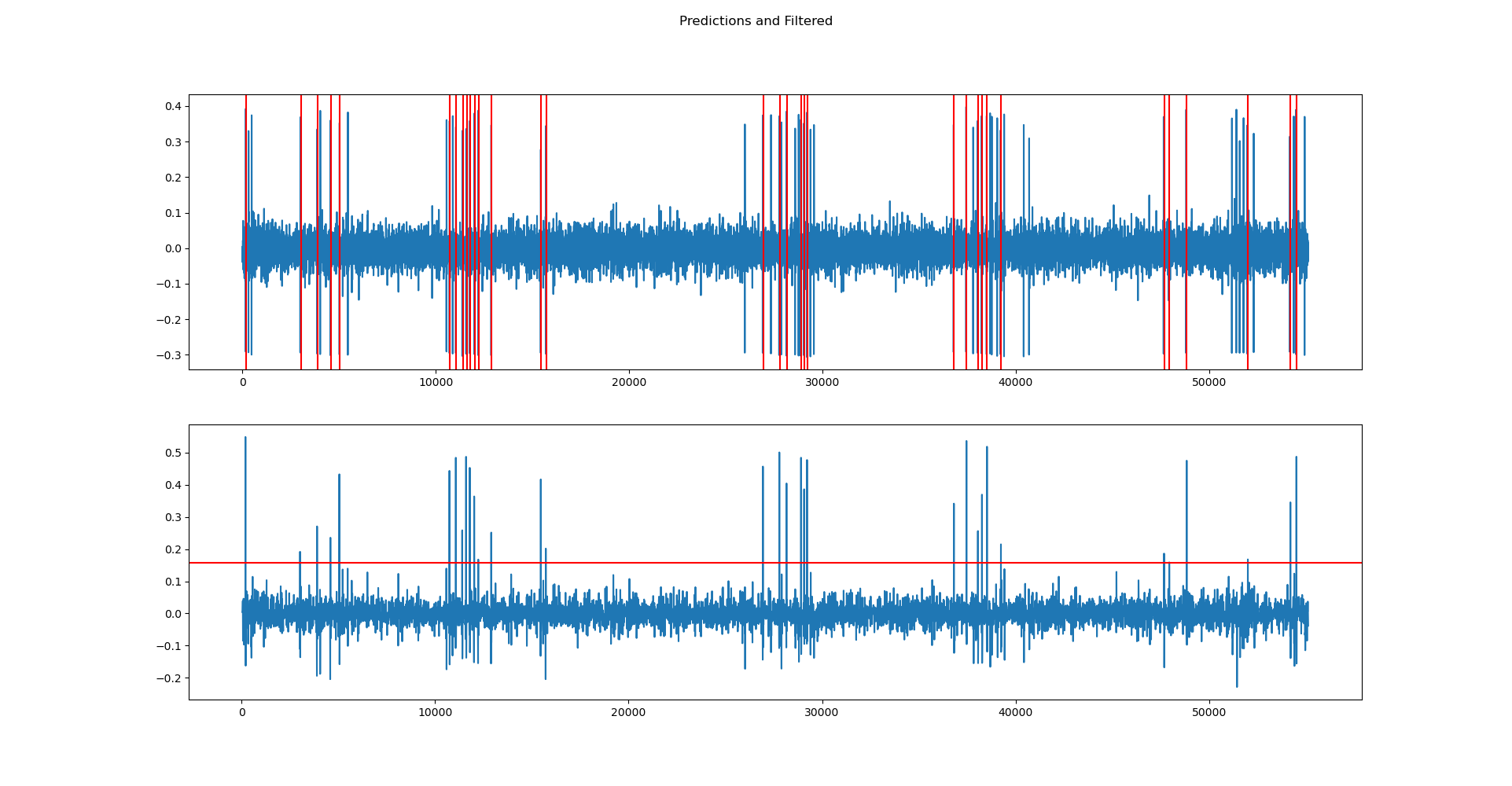
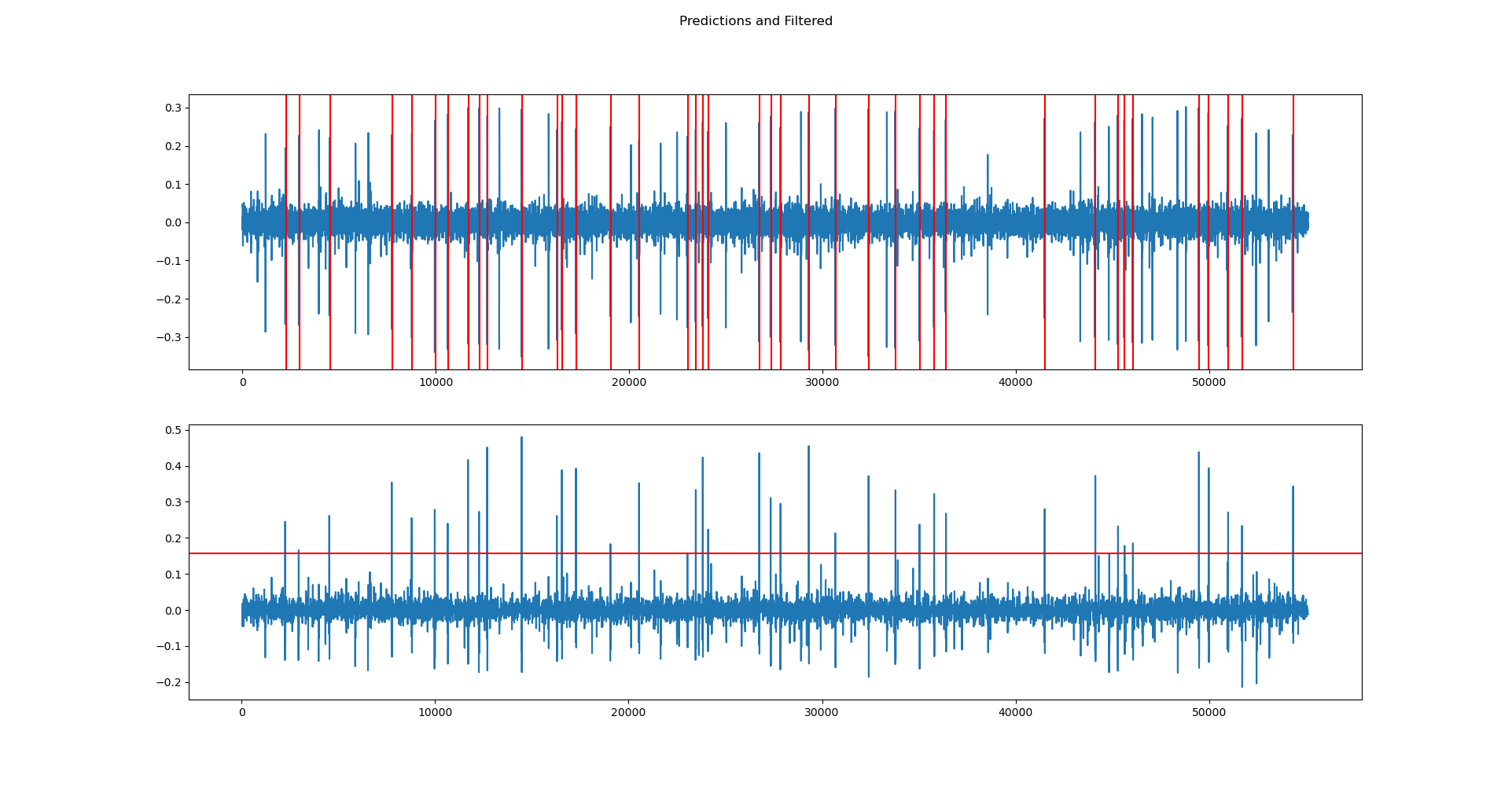
The old assignment from last term, was to find all the tall spikes in the waveform that were overactive parts of the patient’s brain. I accomplished this by using a low pass filter which put all the spikes on a straight line, and then detecting the large spikes above a threshold. One big change I made to the program was to the threshold. Rather than just being a fixed number that I enter in, I found maximum value in the wave, and made the threshold be anything at 60% of its height. That as the spikes change, the threshold would be relative to the individual.

1. 
2. Looking at this plot, we see that there’s some weird distortion between .75ms and 1ms. This was caused by waveform #5(MER0307). I assume this is from some kind of data misalignment.

This is an over lay plot with number 5 removed.

1. The duration of my sample mean is about 1.75 ms, or roughly 40 samples. I picked this number to get it close to the wave form, but not worry about cutting into my mean(in case I wanted to test this on some of the other waves). But made this work variably, so I could change the later if I want.
2. So, to find the standard deviation of the noise, looked at the part of the sample mean that didn’t have the signal(the flat patch between 1.25ms-1.75ms) and used the std function in NumPy, and then too a mean of the array that I got back. I chose this section because it was absent of the signal and had the highest likelihood of being just noise. I ended up getting ~0.024 I tested this by checking the same region in all of the datafiles, and I got approximately the same number, so I assume this was correct.
3. Yes, I think that the noise is pretty close. The shape looks Gaussian and the std looks like it’s in the right place at around ~0.024. I think this makes sense, because the mean should be at zero, and most of the noise visually is 0.024 away from the x axis. This is a histogram of all the noise in all signals.
4. There’s a lot of junk left over in my residuals. I think this has to do more with a misalignment. I have that really gross knee opposite the large opposite spike. I know it’s from recording 5, but I don’t actually know what it is beyond that. I tried removing the file, but I still have a lot of junk on the left. At this point I decided to continue with the program, and save my mean in a csv file, so that I can call it later, without having to recalculate every time. This also allows me to have multiple means and save them for later use. As before, this is my signal with removed number 5, this is what leads me to believe that some of my data points were misaligned. It looks like another wav that was too close.
5. I think it’s reasonable to assume the noise is zero mean gaussian with a ~0.024 std throughout. As stated before, I looked through all of the files to see what I would get, and that’s what I came up with.
6. So the threshold that I chose was based on the Eta variable in Maximum Likelihood. I liked it the most because it moves with the standard deviation and finds the midpoint of the mass of the wave form. I think it seems like the most reasonable threshold to start with. That being said, the gamma in the equation is the log of Theta(the threshold), and I needed to choose my variables. I left my costs pretty basic to start, that way I could see where it lands…. But for my prior probabilities, I tried to figure out what would be best. I decided that I should see how many spikes there were on average from my heuristic. I was going to try with just the annotation data…. But that felt like cheating. So, I used my heuristic, and I found that in the first 10 recordings, there’s an average of about 125 large spikes per recording.

This also led to other questions. I knew for my program I was going to use a sliding window, so I chose my window to be the size of my sample mean 1.75ms, with a 75% overlap(so 40 points, and it shifts over by 10 through each iteration). With 55125 points in the data divided by 10, I look in around 5512 regions, and 125 of those have signals, so my prior probability of detecting a signal is 0.0226(125/55125), and my probability of not finding the signal is ~.9774.

1. For a lot of these, it worked pretty well, except for number 5, which (surprise surprise) was the signal that was creating that gross distortion. For most, it was guessing too much which I could adjust with the threshold, but when I would change my costs to compensate for one of the waveforms, I would greatly mess up number 5, and wouldn’t detect anything.
2. My acceptance interval ended up being the length of my mean. I thought this would keep it simple, and I was hoping that this would help to match exactly what I was looking for. I did notice, that some of the patients, whom had a lot of activity where signals were very close to each other would have many more detections, and this may be partially because I detected more times and got more false positives.

12.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Matched Filter | Sensitivity | Specificity | NPV | PPV |
| MER0102 | 1 | 0.9979 | 0.3443 | 0.6557 |
| MER0106 | 0.9697 | 0.9973 | 0.4576 | 0.5424 |
| MER0109 | 1 | 0.9948 | 0.3537 | 0.6463 |
| MER0119 | 1 | 0.9884 | 0.5442 | 0.4558 |
| MER0303 | 1 | 0.9902 | 0.4901 | 0.5099 |
| MER0307 | 1 | 0.9904 | 0.9798 | 0.0202 |
| MER0308 | 0.9348 | 0.9967 | 0.4342 | 0.5658 |
| MER0309 | 1 | 0.9906 | 0.5053 | 0.4947 |
| MER0313 | 1 | 0.9962 | 0.3725 | 0.6275 |
| MER0601 | 1 | 0.9932 | 0.4964 | 0.5036 |
| ave | 0.99045 | 0.99357 | 0.49781 | 0.50219 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Heuristic | Sensitivity | Specificity | NPV | PPV |
| MER0102 | 1 | 1 | 0 | 1 |
| MER0106 | 1 | 1 | 0 | 1 |
| MER0109 | 1 | 1 | 0 | 1 |
| MER0119 | 1 | 0.9944 | 0.2605 | 0.7395 |
| MER0303 | 1 | 0.9999 | 0.005 | 0.995 |
| MER0307 | 1 | 1 | 0 | 1 |
| MER0308 | 1 | 1 | 0 | 1 |
| MER0309 | 1 | 1 | 0 | 1 |
| MER0313 | 1 | 1 | 0 | 1 |
| MER0601 | 0.9928 | 1 | 0 | 1 |
| ave | 0.99928 | 0.99943 | 0.02655 | 0.97345 |

13. I think that it’s feasible to use the matched filter. I think that my heuristic method, was flawed at returning the correct signals. I think that my signals ended up being misaligned, but if I had spent more time correcting this failure, and having completely aligned signals, I would have had better results.

14. From this homework I learned a lot. After seeing a lot of this used last time, and getting more pointer during the homework presentation, I think I had a better idea how to do this this time around. This homework taught me how much actually goes into making the right matched filter. There’s a lot of preprocessing that needs to be done, and if automated can make things go faster. However, with automation, there’s extra problems that can arise like misaligned data.

15. I’d like to know more about implementation. Last term, you had code samples that you showed us, and I think code samples for just the basic things like implementing likelihood ratio test would be very helpful.