

electronRx Cardio Challenge

Henry Butlin

The task

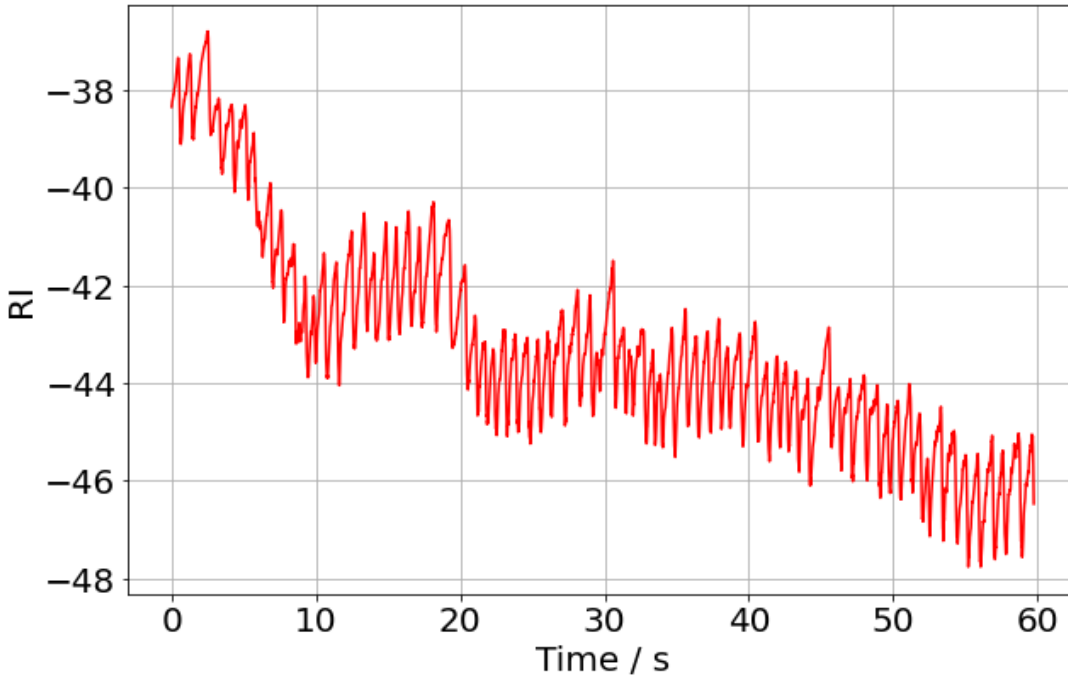
- Signal processing & feature extraction from 10 raw PPG datasets.
 - Cleaning
 - Missing data
 - ...
- Compare to expected results & modify strategy if necessary
- Discriminate SR & AF
- Visualisation



*Thinking process
rather than full code.*

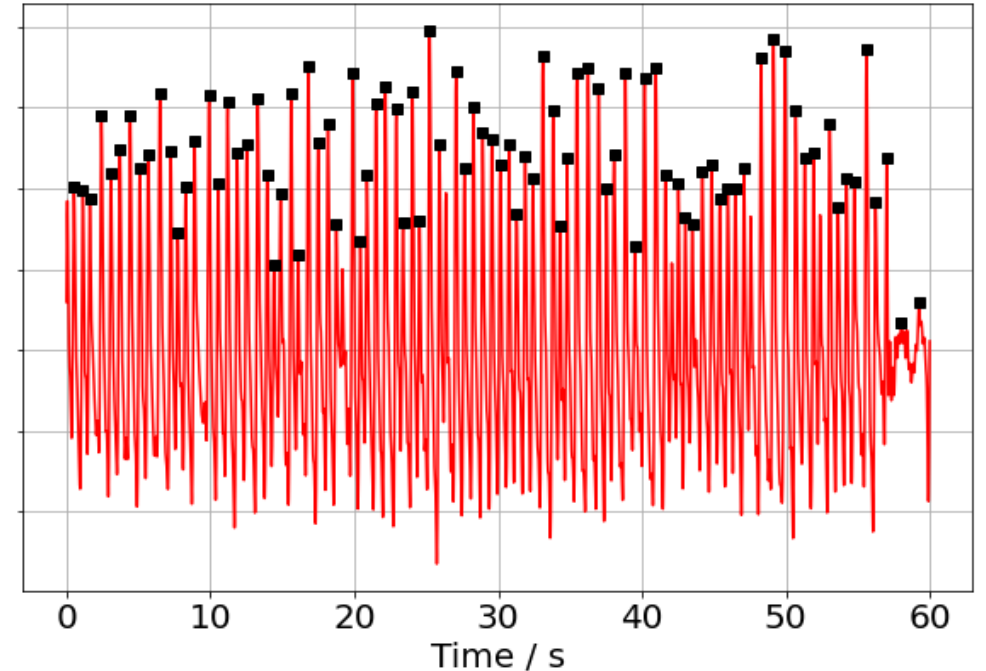
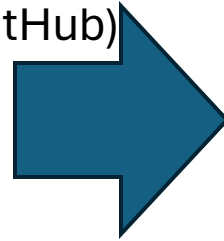
Signal Processing First approach: FT (C-code)

RAW PPG



Noisy, drifting baseline

bandpass filter &
SciPy peak
finding*
(see .c file on
GitHub)



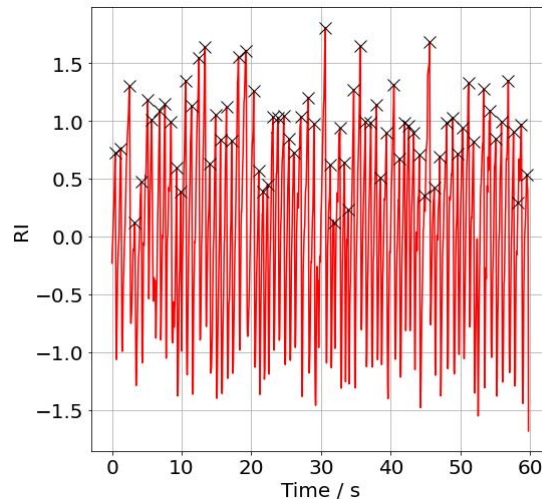
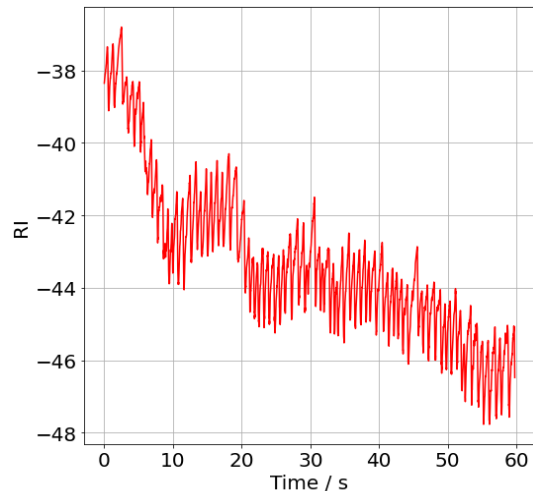
Improved! But:

- Results not a great match to expected
- FT can't handle missing data (see later slide)

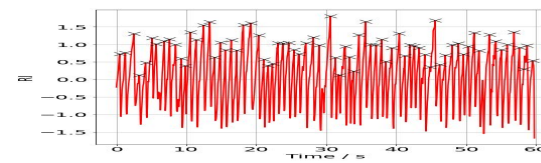
```
* peaks, _ = find_peaks(df.r, height = 0.1, distance = 35)
```

Second approach: Python toolkit

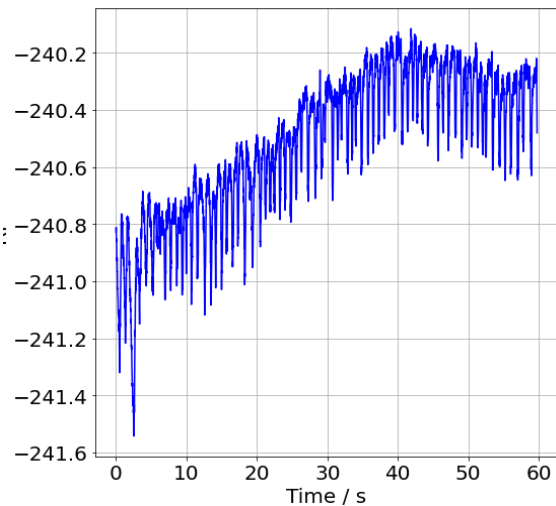
RAW PPG



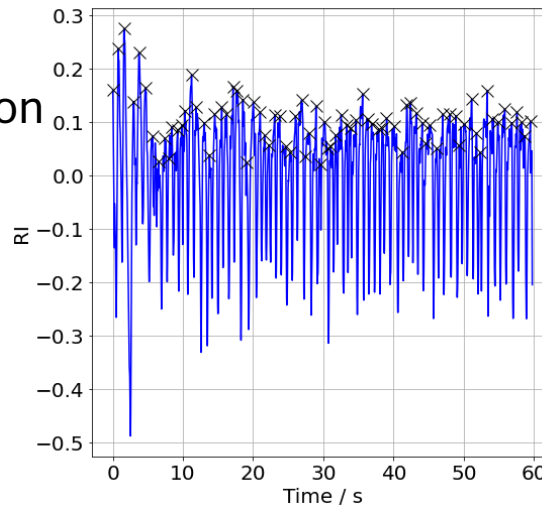
(Rescaled)



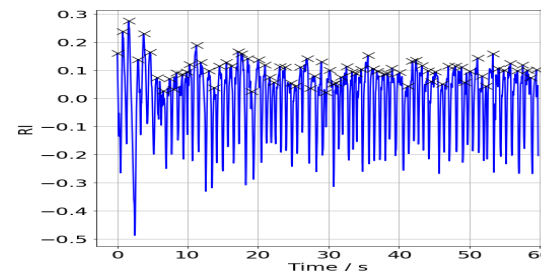
RAW PPG



SciPy Savitzky-Golay noise filter,
polynomial
baseline correction
& peak find



(Rescaled)

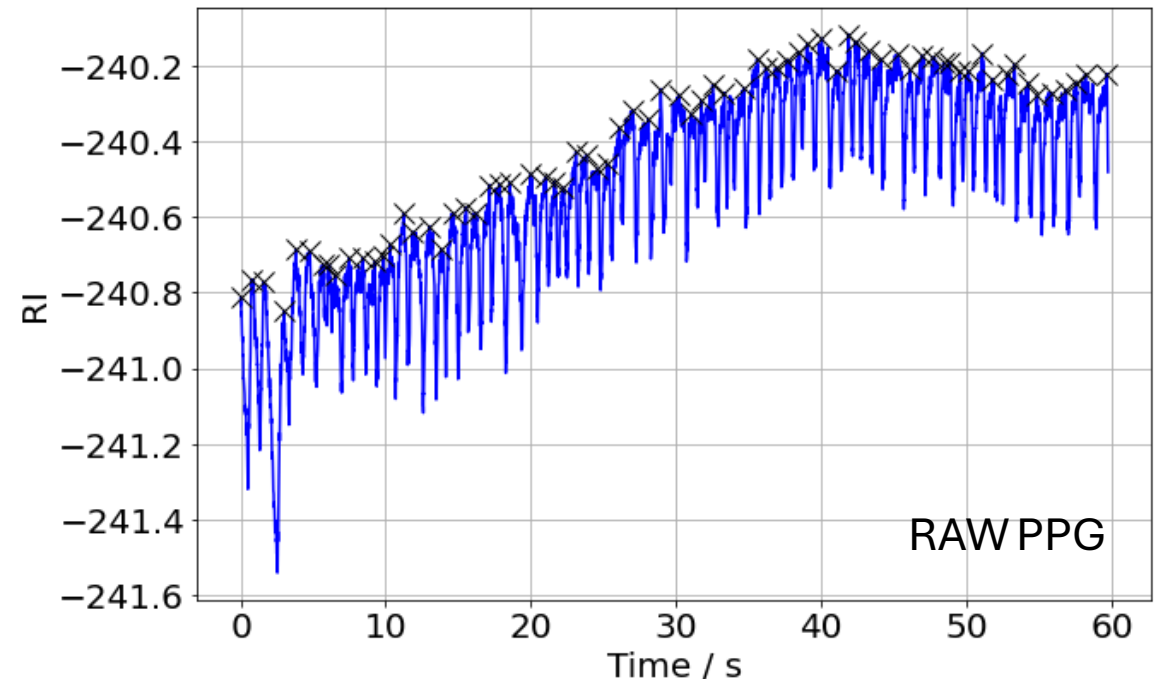
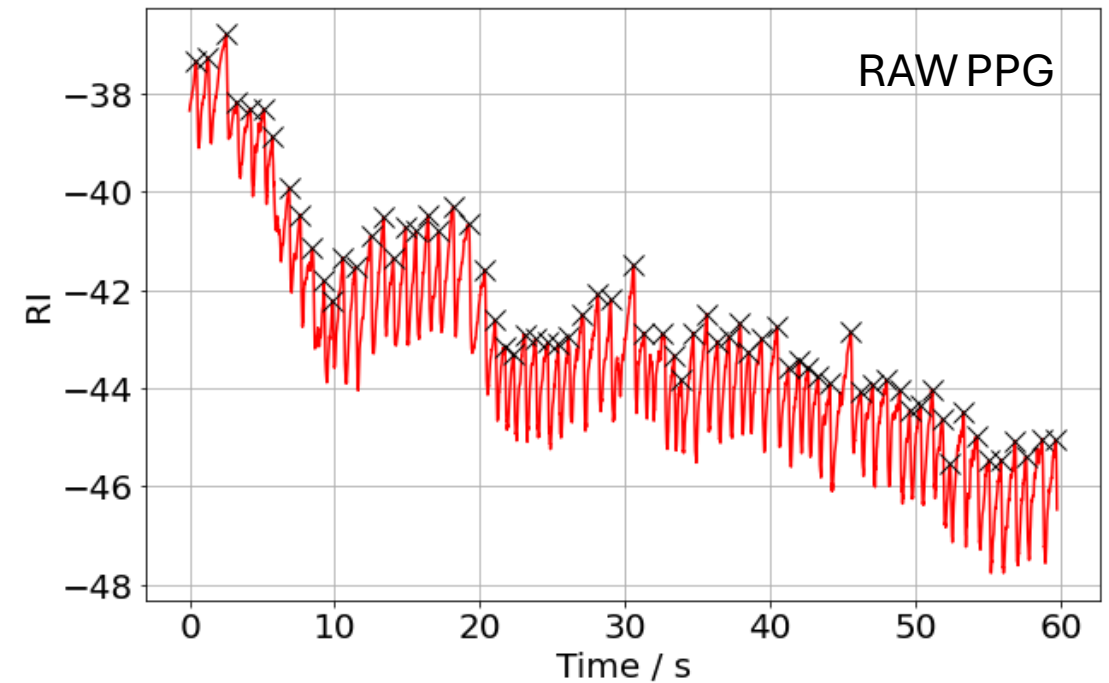


Better!

- Can handle gaps in data
- Results closer to expected

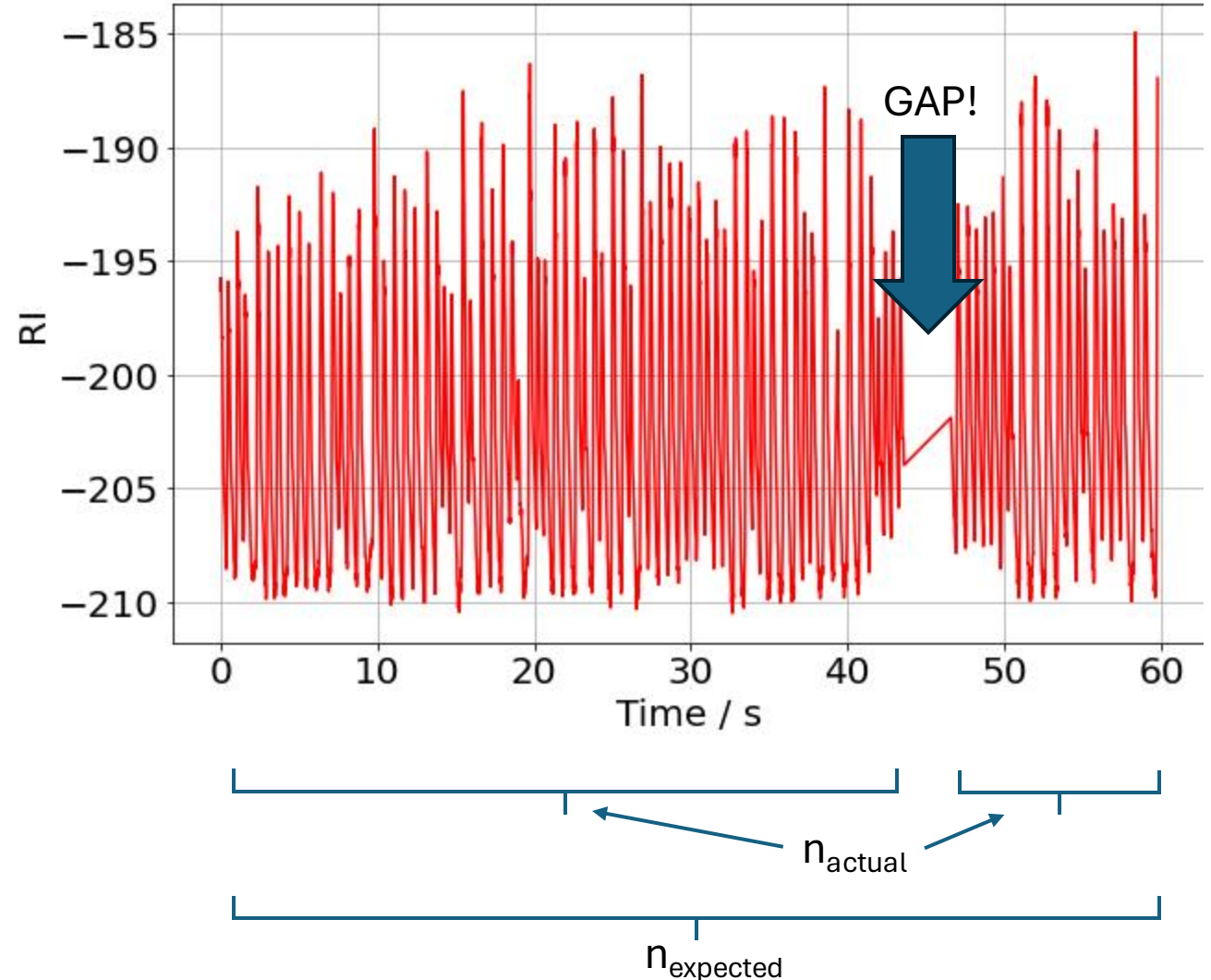
Peak find on raw data?

- To see if the signal preprocessing was having a positive effect, I coded my own peak detection program that didn't need filtered data.
- This worked by finding the positions of the local maxima within incremented windows of data, and yielded visually correct results.



Missing Data

- Some of the raw PPG files were missing rows of data.
- As a quick and crude workaround, obtained heart rate was multiplied by the fraction: $n_{\text{expected}}/n_{\text{actual}}$
- With more time, small gaps could be filled by maximum entropy interpolation, and larger ones by deep learning models such as recurrent neural networks.



Results

- My code yields the rMSSD, SDNN, and pNN50, but for simplicity only HR is displayed here
- In five cases the filtered results are better than the unfiltered ones – my filtering hasn't changed information content, but has visually improved the data
- I investigated the significant outlier in filtered ...

	Expected	Processed	Raw
File	HR (exp)	HR (SG, P)	HR (raw)
2022-06-07 09-15-58	76.0	76.6	77.6
2022-06-07 11-04-55	56.4	56.3	55.3
2022-06-07 11-22-35	89.3	92.0	85.6
2022-06-14 09-42-01	70.6	83.1	70.1
2022-06-07 09-51-55	88.3	87.4	88.4
2022-06-14 09-31-19	54.3	87.0	63.7
2022-06-14 12-55-43	57.1	65.3	52.2
2022-06-14 11-07-24	74.2	79.4	73.3
2022-06-14 11-55-02	66.4	66.1	67.1
2022-06-07 10-03-36	65.4	65.1	63.1

Non-matching Metrics

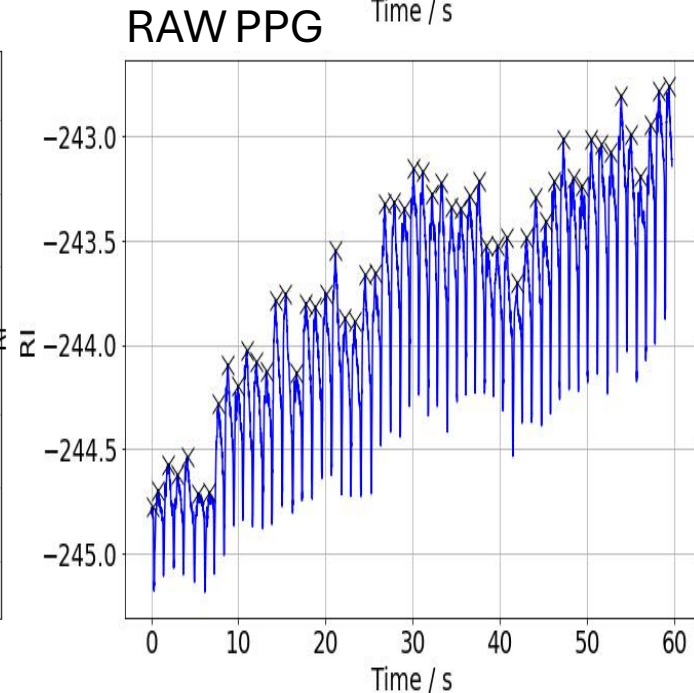
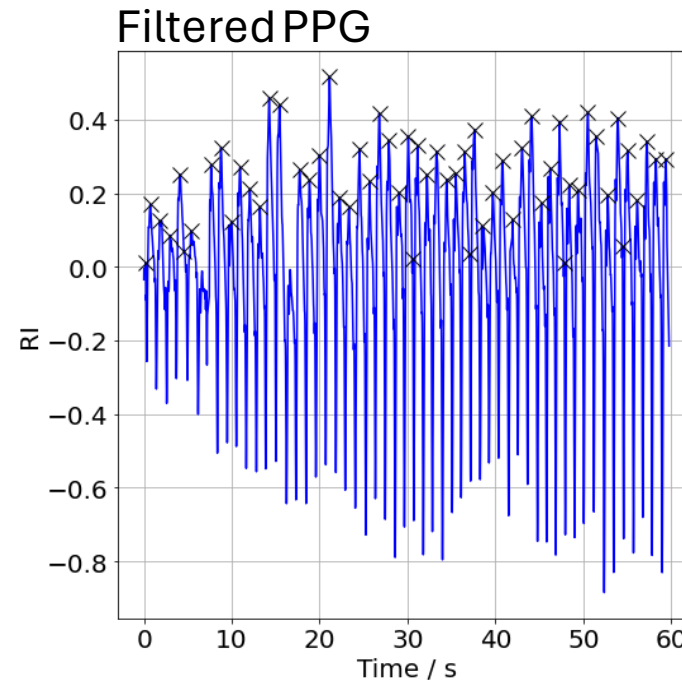
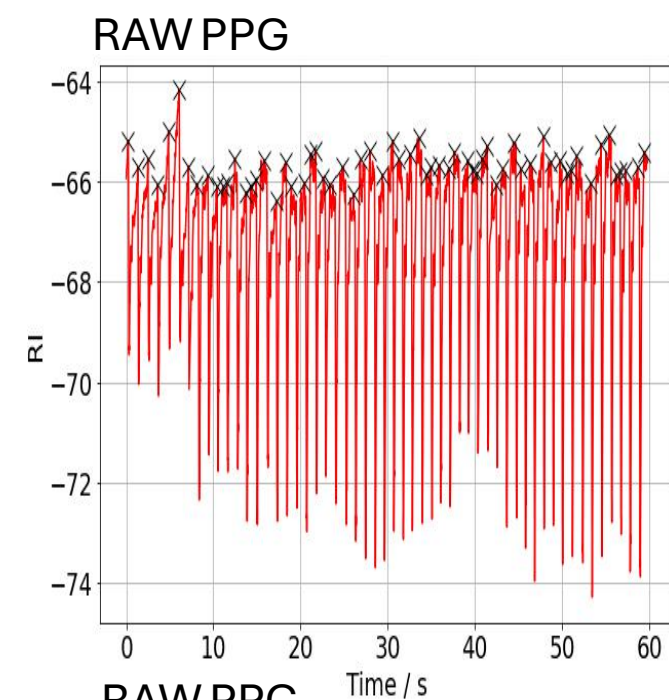
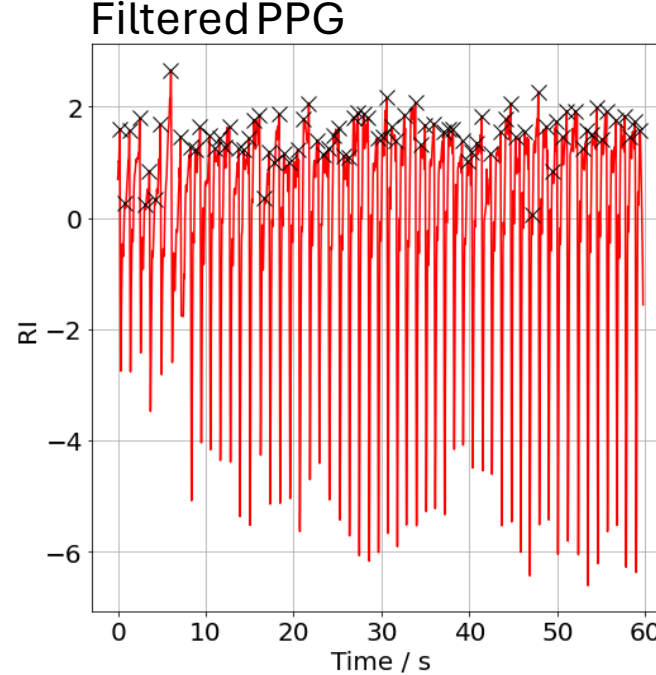
0 / .4

87.0

65.3

- The highlighted anomalous value was the only significantly different value
- The source of this error appears to stem from counting *secondary* peaks
- (May be clinically significant?)
- Could be fixed by a more sophisticated filtering technique or peak detection algorithm (e.g. AMPD*)

*[Article link](#)



Atrial Fibrillation vs Sinus Rhythm

- A study conducted in 2020 found that a deep learning approach was successful in detecting AF from raw PPG data.*
- Here, the rMSSD, SDNN, or pNN50 metrics could be used as crude differentiators between AF and SR.
- As AF is characterised by irregular heartbeats, it seems plausible that a larger range of heartbeat intervals would imply AF.
- A brief visual inspection of the raw PPG waveforms might also be an effective initial screening method.

*[Article link](#)

Visualisation

- Examples of ways to visualise and verify these results.
- In some cases one or more of these plots showed something was going wrong (e.g. peak vs time was non-linear), which would be useful for future debugging.

