ECG in High Intensity Exercise Dataset

Elisabetta De Giovanni, Tomas Teijeiro, David Meier,
Grégoire Millet, David Atienza

The data presented here was extracted from a larger dataset collected through a collaboration between the Embedded Systems Laboratory (ESL) of the Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland and the Institute of Sports Sciences of the University of Lausanne (ISSUL). In this dataset, we report the extracted segments used for an analysis of R peak detection algorithms during high intensity exercise.

1   Protocol of the experiments

The protocol of the experiment was the following.
- 22 subjects performing a cardio-pulmonary maximal exercise test on a cycle ergometer, using a gas mask. A single-lead electrocardiogram (ECG) was measured using the BIOPAC[1] system.
- An initial 3 min of rest were recorded.
- After this baseline, the subjects started cycling at a power of 60W or 90W depending on their fitness level.
- Then, the power of the cycle ergometer was increased by 30W every 3 min till exhaustion (in terms of maximum oxygen uptake or VO2max).
- Finally, physiology experts assessed the so-called ventilatory thresholds and the VO2max based on the pulmonary data (volume of oxygen and CO2).

2   Description of the extracted dataset

The characteristics of the dataset are the following:

- We report only 20 out of 22 subjects that were used for the analysis, because for two subjects the signals were too corrupted or not complete. Specifically, subjects 5 and 12 were discarded.

- The ECG signal was sampled at 500 Hz and then downsampled at 250 Hz. The original ECG signal were measured at maximum 10 mV. Then, they were scaled down by a factor of 1000, hence the data is represented in µV.

- For each subject, 5 segments of 20 s were extracted from the ECG recordings and chosen based on different phases of the maximal exercise test (i.e., before and after the so-called second ventilatory threshold or VT2, before and in the middle of VO2max, and during the recovery after exhaustion) to represent different intensities of physical activity. Fig. 1 shows the segments in time.

  ```
  seg1 --> [VT2-50,VT2-30]
  seg2 --> [VT2+60,VT2+80]
  seg3 --> [VO2max-50,VO2max-30]
  seg4 --> [VO2max-10,VO2max+10]
  seg5 --> [VO2max+60,VO2max+80]
  ```

- The R peak locations were manually annotated in all segments and reviewed by a physician of the Lausanne University Hospital, CHUV. Only segment 5 of subject 9 could not be annotated since there was a problem with the input signal. So, the total number of segments extracted were $20 \cdot 5 - 1 = 99$.
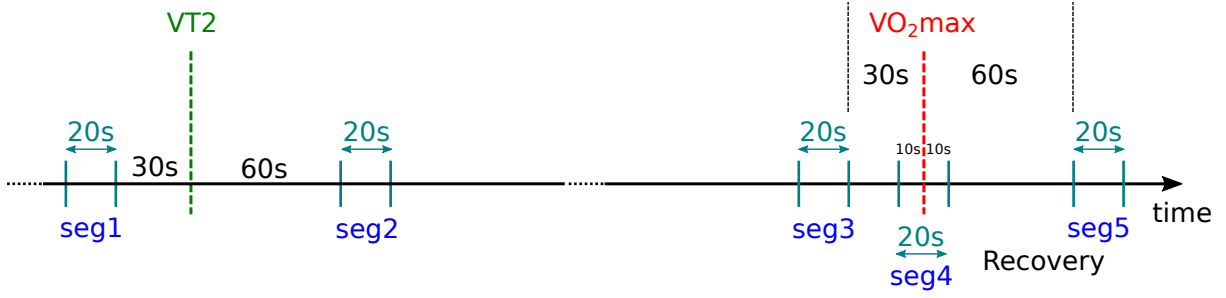
---

[1] <https://www.biopac.com/>

Figure 1: Timeline of segments extracted

3 Format of the extracted dataset

The dataset is divided in two main folders:

- The folder `ecg_segments/` contains the ECG signals saved in two formats, `.csv` and `.mat`. This folder includes both raw (`ecg_raw`) and processed (`ecg`) signals. The processing consists of a morphological filtering[2] and a relative energy non filtering method[3] to enhance the R peaks. The `.csv` files contain only the signal, while the `.mat` files include the signal, the time vector within the maximal stress test, the sampling frequency and the unit of the signal amplitude (µV, as we mentioned before).
- The folder `manual_annotations/` contains the sample indices of the annotated R peaks in `.csv` format. The annotation was done on the processed signals.

[2]<https://infoscience.epfl.ch/record/181118>
[3]<https://infoscience.epfl.ch/record/265401>