Semester Project Suggestion "Child heartbeat detection"

Overview. This project concerns a medical signal processing problem: from electrical signals recorded from a pregnant mother's thorax and abdomen, extract a trace of the unborn child's electrocardial (heartbeat) signal (ECG signal). The difficulty is that the child's ECG signal is very weak and almost entirely hidden under the much stronger ECG and other electrical signals that come from the mother.

Difficulty. This is a nontrivial and clinically relevant task, and all sorts of basic and advanced signal processing and machine learning methods have been investigated for tackling it. A survey article is

R. Kahankova et al., *A Review of Signal Processing Techniques for Non-Invasive Fetal Electrocardiography*. IEEE Reviews in Biomedical Engineering, August 2019, DOI: 10.1109/RBME.2019.2938061 (online copy at https://www.researchgate.net/publication/335500003_A_Review_of_Signal_Processing_Techniques_for_Non-Invasive_Fetal_Electrocardiography also attached to the project materials)

While there is no upper limit in sophistication and model complexity which one might invest, halfway good-looking solutions can already be obtained with elementary linear-regression based methods, combined with likewise elementary high-pass filtering for signal preprocessing. If you have an idea what "high-pass filtering" means this project might be to your taste (and it is doable in the few weeks that you have); if no member in your team has an idea of what "high-pass filtering" means, this project has likely too much signal processing flavor added to be doable without too much additional self-study in signal processing.

Data source. A number of benchmark datasets are indicated in the survey article mentioned above. However, I suggest to use a relatively lightweight dataset which I took from a similar academic semester project from Lee Cheng, originally (in the year 2016) found at http://www.masys.url.tw/AU/AU.htm and http://www.masys.url.tw/AU/2015SP/BMSD-D/HW/HWfinal-Maternal_Fetal_ECG/ProjectDescription.htm. These pages are no longer reachable. Lee Cheng's dataset consists in 5 ECG channels "thorax1", "thorax2", "abdomen1", "abdomen2", "abdomen3", each consisting of 20,000 sample points recorded at a sampling frequency of 1000 Hz, thus the total recording time is 20 sec. Figure 1 shows a 2,000 step segment of these five traces (normalized to unit variance and zero mean).

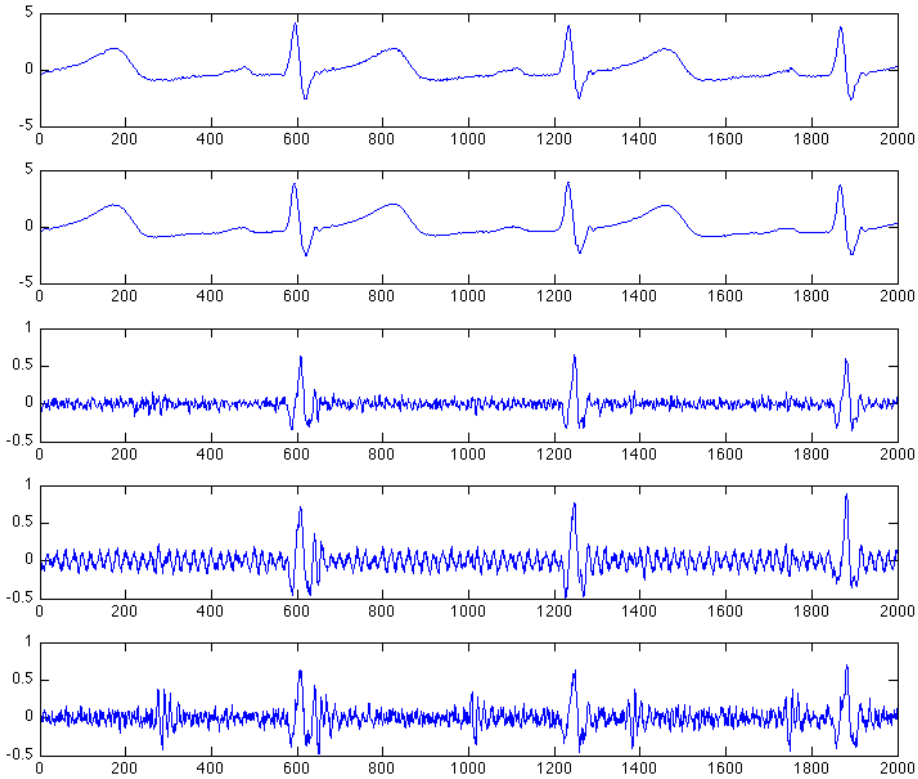

Figure 1. For explanation see text

A trace of the child's heartbeat ECG can be best seen in the last of these channels – the child's heartbeat is about twice as fast the mother's.

Figure 2 shows where these five channels have been recorded.

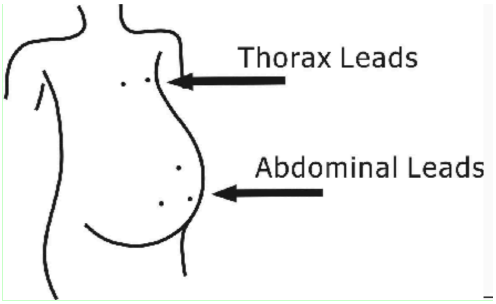

Figure 2: locations of 5 electrodes. Image taken the legacy webpage of Lee Cheng.

I packed the five traces in five .txt files, included in the materials for this project (file "ECGdata.zip").

Task: Learn a filtering algorithm which, if it gets the five mother ECG signals as inputs, generates an output signal which shows as much of the child's ECG and as little of the mother's ECG as possible. Note that it is not possible to extract the "real" child ECG, because this electrical signal will have undergone unknown transformations before reaching the recording electrodes. If you manage to filter out a

signal which looks halfway clean and has some sorts of "pulses" at a rate roughly twice the mother's heartbeat rate, then you are doing fine.
Figure 3 shows what I could achieve using only linear methods:

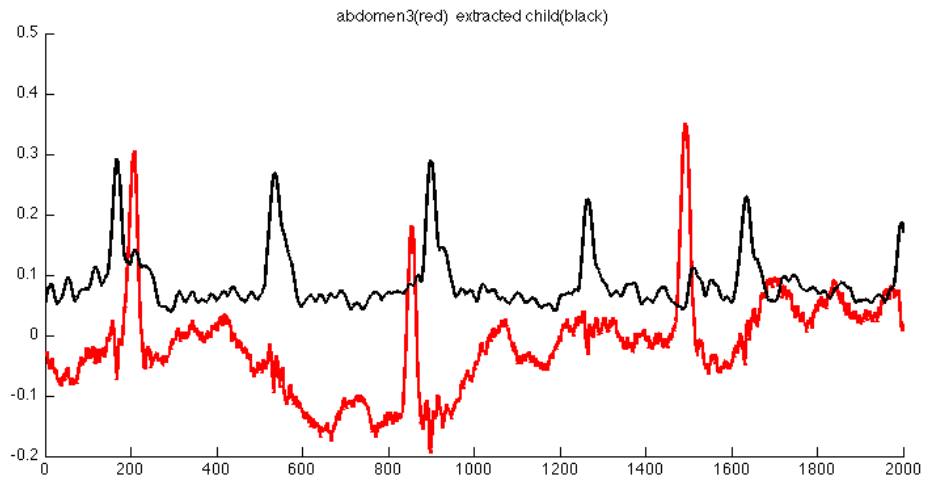

Figure 3: My private first stab at this task. The black line is the output of my filter, the red line is one of the five inputs.

Approach / basic architecture. There is a classical approach to solve this kind of signal processing tasks – that is, given several recordings, all dominated by a signal source that one wants to cancel, and a few of them showing a faint trace of a signal that one wants to bring to the foreground. This task is called *noise cancellation* (here the "noise" is the mother's ECG). It is a classical task and there is a classical approach to solve it. The core idea is to use input channels that have (almost) no child ECG component (the thorax recordings) to *predict* the channel which has the strongest child component (that is "abdomen3"), then subtract that prediction from abdomen3 – that is, subtract everything that can be explained from mother ECG input. With some additional pre- and postprocessing (smoothing, mostly), I got what you see in Figure 3. This classical approach of *noise cancellation* is outlined in our ML lecture notes Section 11.2.3. There, the principle of this approach is explained for online adaptive filters, but the same principle can also be used with non-adaptive filters that are trained once and for all on a given training data set – which is what you would be probably doing.

Hint. As you see in Figure 3, the ECG signals have a moving shift in baseline level. This is an effect that will derail most standard machine learning algorithms. I very much recommend to first do a high-pass filtering on the signals which removes this (low-frequency!) wandering baseline.