B18 Biomedical Modelling and Monitoring (A10589)
Wearables Laboratory

Dr. Andrew P. Creagh
andrew.creagh@eng.ox.ac.uk

Department of Engineering Science
University of Oxford

Hilary Term 2022

Last Modified: ..... January 11, 2022
Organiser(s): ..... Prof. David A. Clifton
Instructor(s): ..... Dr. Andrew P. Creagh

Contents

1 Lab Overview

The aim of this laboratory is to gain hand on experience with various different biomedical signals that can be collected from wearable sensors: electrocardiogram (ECG), electroencephalogram (EEG) and smartphone accelerometry. You will learn how to acquire, pre-process and analyse various sensor-based measurements from clinically recorded data. The goal of this lab is to understand the role of signal conditioning, how to choose suitable parameters of signal processing algorithms, and the various technological approaches biomedical engineers have to understand diseases from wearable sensor data. It is recommended to follow to pdf version of this lab sheet to enable clickable embedded hyperlinks, depicted in **blue**.

2 Getting Started

This laboratory will require a MATLAB vR2019b (or later, for example vR2020a) installation on the machine you are using (The MathWorks, Natick, MA, USA). To download a version of MATLAB on your own machine see; `https://help.it.ox.ac.uk/sls/fulllist`. Follow instructions for download, using the licence-key provided by the ENG department. More details on MATLAB installations can be found in the MATLAB documentation. Ensure you have installed the signal processing toolbox `https://uk.mathworks.com/products/signal.html`.

2.1 Downloading Data

Once MATLAB is installed you must next download the data and accompanying code for the lab. The laboratory materials, scripts and data can be downloaded from: `https://github.com/apcreagh/B18-Wearables-Laboratory`. Download the lab files and unzip the content to your ***own*** MATLAB home directory (this is usually somewhere like: `'~/user/Documents/MATLAB/'`). Type `userpath` into the command window to find this. Next run the `start_up.m` file to extract and unzip the data and files needed for the B18 lab. It is recommended that you store all data in the `'/DATA/'` folder which has been automatically created and all scripts and code in the `'/SCRIPTS/'` folder which has been automatically created. Remember to ensure that you use the correct `pathname` when loading your data, as this is the biggest troubleshooting problem students typically encounter. See the official MATLAB documentation for further useful information.

2.2 Submission

The B18 lab should be implemented using MATLAB. You are required to submit a laboratory report as a digital file (for example a .doc, .pdf, or similar file), which includes your preparatory worka answers, as well as your completed exercises. It is recommended you use Microsoft Word, Microsoft OneNote, LaTeX, or similar software to complete your lab sheet exercises. Your should cite references where used in any of your answers. Students are expected to complete all exercises and document results in their lab write-up, including any figures and tables generated. All figures should have their axis (and units) labelled where applicable. Figures and tables should be saved an inserted into documents, or can be copy and pasted into a document. Details on how to copy figures to clipboards can be found in the MATLAB documentation. Please label all exercises clearly. Failure to clearly document your answers and discussions may result in a loss of marks. Students should submit their accompanying MATLAB scripts, which should be clearly labelled. It is recommended you make use of code sections to help with layouts and formatting, see MATLAB documentation.

3 Background

3.1 The Electrocardiogram

The electrocardiogram (ECG) records the electrical activity generated by the heart as it undergoes depolarisation and repolarisation of the atria and ventricles. The electrical currents that are generated from this process propagate through the body. A typical ECG waveform is shown in figure 1.

3.1.1 Recording ECG

There are multiple different approaches to measure ECG, with varying degrees of precision, viewpoint, and localisation by placing electrodes at different positions on the human body. The concept behind ECG is based on Einthoven's triangle, an imaginary formation of three limb leads in an equilateral triangle with the heart at the centre, which forms a reference system to analyse ECG waveforms. There are three standard lead placements forming the *bipolar* standard limb leads: lead I, an axis from left arm (+ electrode) to right arm (− electrode); lead II, axis goes from the right arm (−) to the left leg (+); and lead III, an axis from the left arm (−) to the left leg (+), as depicted in figure 2a. Additionally there are three augmented *unipolar* limb leads which are single positive electrodes that are referenced against a combination of the other limb electrodes. These leads can be "augmented" or constructed from the *bipolar* leads, allowing us to use same electrode placement as the *bipolar* form. The positive electrodes for these augmented leads are located on the left arm (aVL), the right arm (aVR), and the left leg (aVF), as shown in figures 2c. The combination of the 6 *bipolar* and *unipolar* leads records electrical activity along a single plane, termed the frontal plane relative to the heart (see figure 3a). There are also six precordial, unipolar chest leads which are sometimes applied in rigorous clinical settings. These places six positive electrodes on the surface of the chest over different regions of the heart in order to record electrical activity in a plane perpendicular to the frontal plane (see figure 3b).

Following your lecture notes, diagnostic information can be obtained from the ECG waveform, by analysis of the amplitude and relative timing of the various segments. EEG can be used for ambulatory monitoring, exercise and stress analysis or even for foetal ECG monitoring. Informative ECG features such as the heart rate, the timing of the PQRST complexes or heart rate variability (HRV).

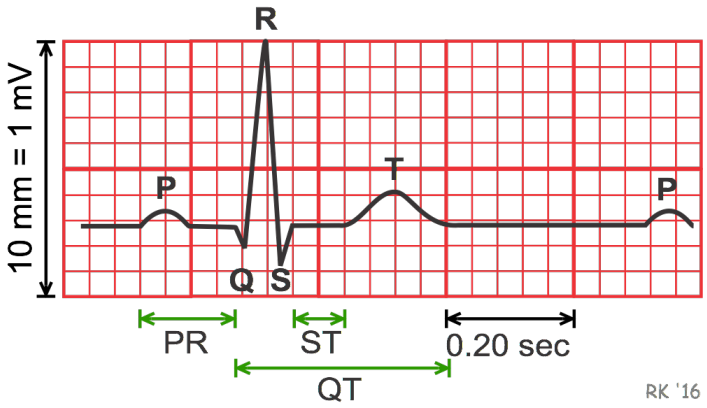
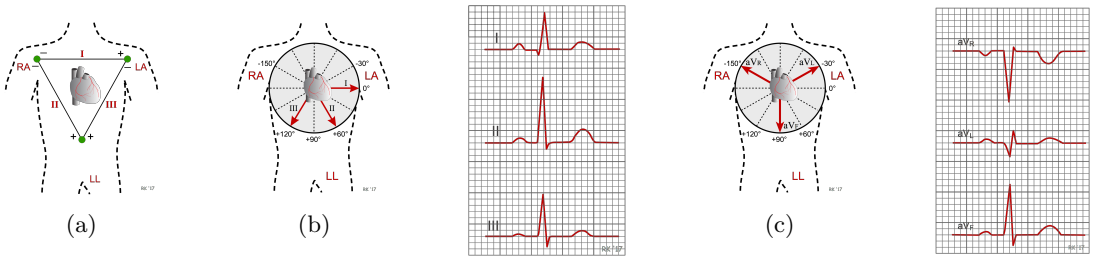
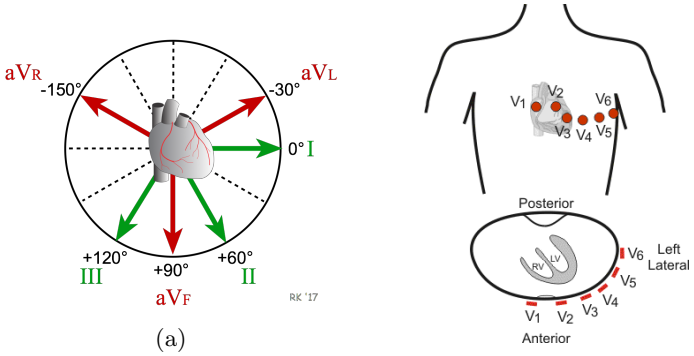


Figure 1: The typical ECG waveform. Image credit: http://www.cvphysiology.com/Arrhythmias/A009.htm



(a) (b) (c)

Figure 2: Electrode placement for (a) the standard *bipolar* 3-lead configuration and Einthoven's triangle construction; (b) *bipolar* limb lead axis and resulting ECG trace; (c) *unipolar* augmented 3-lead axis and resulting ECG trace. Image credit: http://www.cvphysiology.com/Arrhythmias/A013a.htm



(a) (b)

Figure 3: A full 12-lead ECG axis: (a) the combined *bipolar* and *unipolar* 6-axis ECG axis; (b) the ECG chest leads placement. Image credit: http://www.cvphysiology.com/Arrhythmias/A013c.htm

3.1.2 Heart Defects and Diseases

As you've learned in the lecture notes, the heart beat rhythm is governed by pacemaker cells within the sinoatrial (SA) node which dictates the depolarisation and repolarisation of the atria and ventricles. Unfortunately, when this rhythm becomes irregular, too fast (tachycardia) or too slow (bradycardia), or the frequency of the atrial and ventricular beats are different, arrhythmias occur. Atrial fibrillation is one of the most common rhythm disturbances where this condition causes an irregular and often abnormally fast heart rate. Atrial fibrillation (AF) affects around 1 million people in the UK, affecting about 7 in 100 people aged over 65. Patients may describe an arrhythmia as a palpitation or fluttering sensation in the chest. Luckily, many arrhythmias can be treated with supression antiarrhythmic drugs. It is therefore critical to identify arrhythmias, such as AF, as early as possible so that patients can be provided with vital treatments.

Other common heart defects and diseases include cardiac valve diseases, coronary artery diseases, hypertension and hypotension.

3.2 The Electroencephalogram

The electroencephalogram (EEG) is a recording of brain activity; the summation of neural depolarisations in the brain due to stimuli from the five senses as well as from normal brain activity. On the surface of the brain, these potentials are of the order of 10 mV, however the amplitude of the EEG signal recorded with scalp electrodes is only ever 100 μV at most because of the attenuation caused

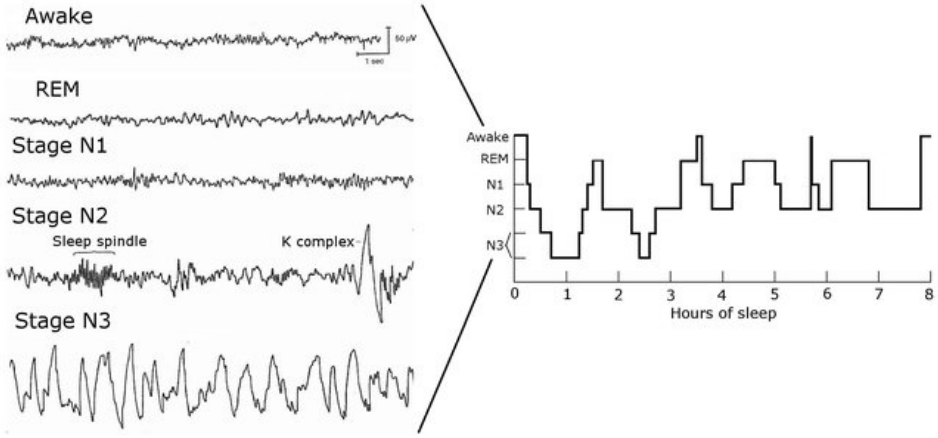


Figure 4: Typical EEG traces of the different stages of adult sleep. On the right is a typical hypnogram, which plots the stage of sleep over time. Note how slow wave sleep (Stage N3) progressively decreases in duration throughout the night, while REM sleep increases in duration. Credit: [1].

by the skull.

The clinical and experimental uses of EEG are wide ranging, such as investigating epileptic fits, studying sleep, the diagnosis of brain death, to language and visual processing. The use of EEG for sleep staging is presented in figure 4. There are four sleep stages in adult sleep: one for rapid eye movement (REM) sleep and three that form non-REM (NREM) sleep. NREM sleep consists of N1, N2, and N3 (or slow wave, deep sleep).

The frequency content of the EEG varies with the state of alertness and mental activity. To assist in EEG analysis, the normal EEG range of 0.5 to 30 Hz has been sub-divided into five bands (note, however, that there is some degree of variation in the exact cut-offs from one system to another:

- Delta δ (0.5–4 Hz)
- Theta θ (4–8 Hz)
- Alpha α (8–13 Hz)
- Beta β (13–22 Hz)
- Gamma γ (22–30 Hz)

3.3 Physical Activity Monitoring

Human mobility and physical activity patterns are well-established indicators of health and quality of life [2, 3, 4].

The rapid evolution of wearable sensors and devices for the collection of health-related data offer the opportunity to monitor mobility, gait and daily-physical activity patterns in patients. Nowadays, consumer wearable sensors, such as smartphone and smartwatches specifically, are ubiquitous in everyday life. Often termed inertial measurement units (IMU), or inertial sensors, these wearable devices generally contain 3-axis accelerometer, gyroscope, and magnetometer sensors. Figure 5 demonstrates the smartphone and smartwatch device inertial sensor coordinates.

Many studies have demonstrated the benefits of wearable-based daily activity monitoring to indicate

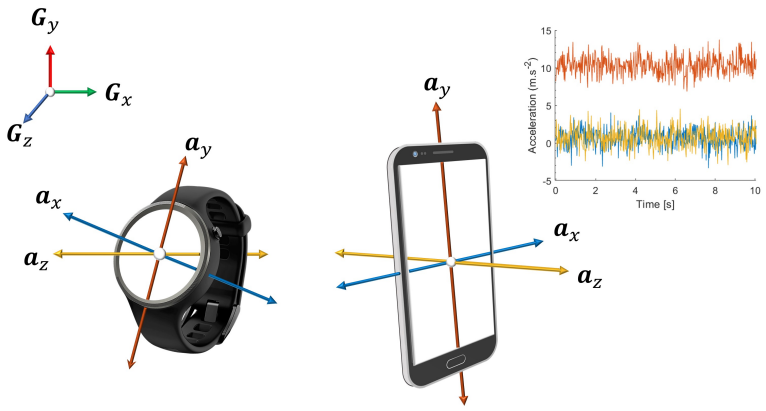


Figure 5: Smartphone and smartwatch device inertial sensor coordinates. Both devices contain a 3-axis accelerometer sensor, $(\boldsymbol{a}_x, \boldsymbol{a}_y, \boldsymbol{a}_z)$, which at any instantaneous time point t is incident to the direction of gravity (vertical), defined globally as $(\boldsymbol{G}_x = 0, \boldsymbol{G}_y = -9.81\ m.s^{-2}, \boldsymbol{G}_z = 0)$. Image credit: [7].

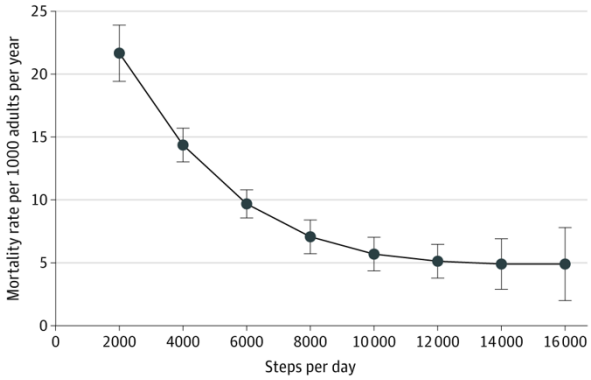


Figure 6: The association of daily step count with mortality rate amoung US adults at least 40 years old (n=4840). Credit: [5]

health stats, quality of life and to monitor patients. For example, it has been shown that a greater number of daily steps was significantly associated with lower mortality rates in a representative sample of US adults [5]. Other studies have shown that physical activity decreases with age in a representative sample of UK adults [4]. Furthermore, a wide range of diseases have a negative effect on physical activity, affecting the amount, type and way that patients perform certain activities and tasks [6].

4 Section A: The Electrocardiogram

This section aimed to introduce some standard approaches to analysing electrocardiogram (ECG) data. This example exercise uses clinical ECG data from the PhysioNet 2017 Challenge [8]. PhysioNet (a the moniker of the Research Resource for Complex Physiologic Signals) is a community resource and archive of digital recordings of physiologic signals, time series, and related data for use by the biomedical research community.

The 2017 PhysioNet/CinC Challenge aimed to encourage the development of algorithms to classify, from a single short ECG lead recording (between 30 seconds and 60 seconds in length and all signals are sampled at 300 Hz), whether the recording shows normal sinus rhythm, atrial fibrillation (AF), an alternative rhythm, or is too noisy to be classified. More details on the PhysioNet 2017 Challenge can be found at `https://physionet.org/content/challenge-2017/1.0.0/`. The data consists of a set of ECG signals sampled at 300 Hertz (Hz) and divided by a group of experts into four different ECG classes: Normal (N), AFib (A), Other Rhythm (O), and Noisy Recording (∼).

4.1 Exercise 1: Exploratory ECG Data Analysis

This section begins the lab by exploring the data in the PhysioNet challenge. In this exercise we will plot a number of ECG examples from various subjects and recordings for different healthy (normal) and abnormal (AF, or other) examples.

You have been provided with example script to perform the data loading, data wrangling, and plotting utilities, needed to complete this section. See the `B18_ECG_SectionA_Excercise_1.m` file. **Choose any recordings you want from the whole PhysioNet data for all subsequent analysis in this exercise. Do NOT use the same example recordings in this lab sheet, namely: A00007, A00267, A00077, A01246**.

1. Load the PhysioNet data into MATLAB;
2. Plot a suitable portion of ECG data from a normal subject, which depicts at least 2 heartbeats. Identify and label the PQRST points;
3. Visualise a segment of one ECG signal from each of the 4 classes from any chosen set of examples. Label each recording. See figure 8 for an example;
   (a) Describe some of the signal characteristics observed that distinguish normal ECG versus AF ECG in your chosen examples. Use extra plots to illustrate your discussion points if required;
   (b) With respect to your chosen noisy ECG example, or any artefacts observed in other examples, briefly discuss some of the reasons why ECG signals can be corrupted or noisy.

4.1.1 Hints:

- The proportion of Normal (N), AFib (A), Other Rhythm (O), and Noisy Recording (∼) classes in the dataset can be determined from the `RecordingInfo` table, using the `tabulate.m` function;
- The `text.m` function (see MATLAB documentation for some examples) can be useful for adding annotations to plots directly in MATLAB.
- To convert samples to time, simply create a separate monotonic vector of samples, converting samples to time as we all know $f = 1/t$:

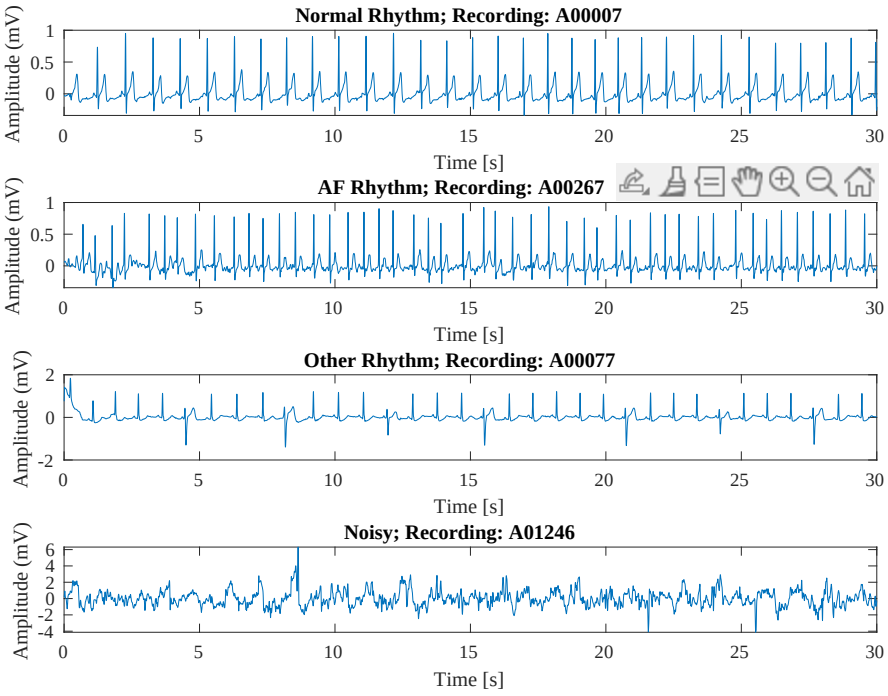


Figure 7: Examples of ECG waveforms for a normal rhythm, AF, other rhythm and a noisy recording.

```
%FYI: the PhysioNet data is sampled at 300 Hz.
fs=300; %[user set variable]: Sampling frequency (Hz)
samples=1:length(signal); % 'signal' is your vector of ecg values
%'time' is a corresponding monotonic vector of time values
time=samples./fs;
```

4.2 Exercise 2: Pre-Processing ECG

Signal conditioning and pre-processing are important steps to clean and prepare data for further analysis later. Sometimes ECG signals contain an underlying trend, such as a change in the mean over time. Detrending is therefore applied by removing (subtracting) the mean component from the signal. Typically, the signal will be "normalised" or "standardised" to have unit variance. This standardisation is often performed using the z-score. For sample data $\mathbf{x}$ with mean $\overline{\mathbf{x}}$ and standard deviation σ, the z-score of a data point x_i is:

$$z_i = \frac{x_i - \overline{\mathbf{x}}}{\sigma} \tag{1}$$

where the sample standard deviation σ is given by:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \overline{\mathbf{x}})^2}{N-1}} \tag{2}$$

Filtering ECG signals appropriately becomes the next step in pre-processing. Often signals are filtered with a combination of low-pass and high-pass filters, i.e. band-pass filters. You have been provided with a separate function to perform filtering `filter_ecg.m` using a Butterworth IIR band-pass. Butterworth

filters have a smooth monotonic frequency response that is maximally flat in the passband, whereby they sacrifice rolloff steepness in favour of flatness. Using the normal and AF recording examples identified as part of exercise A.13:

1. Detrend the data by subtracting the mean component; then, standardise the detrended data using the z-score approach;
2. Using the same standardised and detrended normal and AF recording examples above, apply band-pass filtering to each recording (after detrending) using the `filter_ecg.m` file provided;
   (a) Identify suitable filtering parameters: high-pass and low-pass frequency thresholds, filter order;
      i. Justify and discuss your choice of (1) high-pass frequency threshold; (2) low-pass frequency threshold with respect to your understanding of ECG signals;
3. Plot the resulting recordings for both your chosen normal and AF examples after detrending, standardisation and filtering has been applied.

4.2.1 Hints:

- MATLAB have in-build functions for detrending (`detrend`) and normalising (`zscore`)!
- You can easily implement detrending and normalisation yourselves. Remember, using element-wise operators in MATLAB reduces the need to construct `for` loops. For example the operation `x./y` divides each element of `x` by the corresponding element of `y`. See MATLAB documentation for more details.
- The `filter_ecg.m` function useage is as follows:

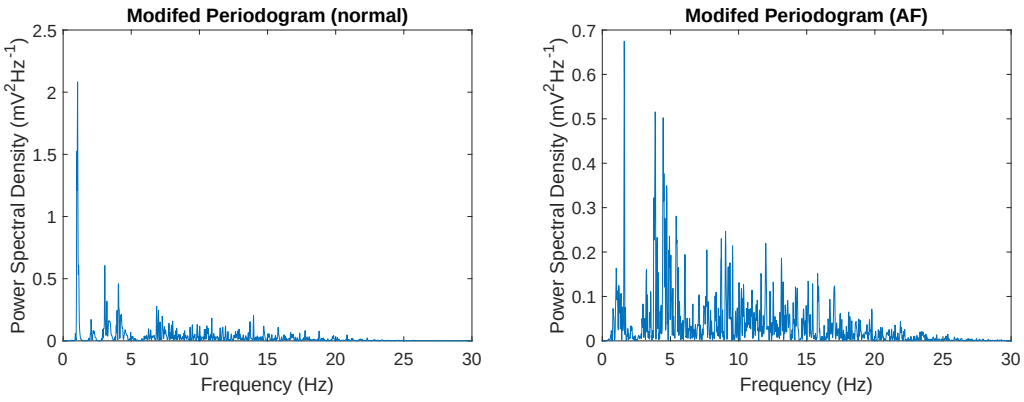
```
org_signal; %[user set variable]: the original unfiltered ECG
    signal
fs=300;     %[user set variable]: sampling frequency  [Hz]
f_low=100;  %[user set variable]: low-pass frequency  [Hz]
f_high=0.5; %[user set variable]: high-pass frequency [Hz]
norder=4;   %[user set variable]: filter order

%filter_ecg.m filters the orginal signal and outputs
%the band-pass filtered ECG signal
filtered_signal=filter_ecg(org_signal, fs, f_high, f_low, norder);
```

- This is a question students can get stuck on. Consider what make sensible filtering parameters: we want to filter out high-frequency noise in the signal (for example, 50 Hz from A/C power supply) and low-frequency noise (for example breathing <0.5 Hz), yet retain important frequency components in the ECG (for example the HEART RATE! usually at around 60 BPM $\approx$ 1 Hz).

4.3 Exercise 3: Frequency Representation of ECG Signals

A periodogram can be used to determine the frequency representation of an ECG. The periodogram is a non-parametric estimate of the power spectral density (PSD) of a signal which uses discrete Fourier transform (DFT) through the fast Fourier transform (FFT) algorithm. The PSD estimate of ECG vector $\mathbf{x} \in \mathbb{R}^{1\times N}$ can be performed using the modified periodogram. The modified periodogram multiplies the input time series by a window function, in this case with a Hamming window function



(a) (b)

Figure 8: Modified periodogram power spectral density (PSD) estimation of (a) a normal ECG recording (b) an AF ECG recording.

h_n. The modified periodogram PSD is defined by:

$$\widehat{P}(f) = \frac{\Delta t}{N} \left| \sum_{n=0}^{N-1} h_n x_n e^{-j2\pi fn} \right|^2 \tag{3}$$

where the periodogram is the DFT of the biased estimate of the autocorrelation sequence for the n^{th} sample x_n, sampled at f_s per unit time; Δt is the sampling interval.

The periodogram is not a consistent estimator of the true PSD however. The Welch's Overlapped Segment Averaging Spectral Estimation aims to reduce the variance of the periodogram by segmenting the time series into overlapping segments and computing the modified periodogram for each segment. These PSD segments can then averaged give a more consistent estimate of the power spectral density.

Using the same standardised, detrended and band-pass filtered data from normal and AF recording examples in exercise A.2.3:

1. Plot the modified periodogram PSD estimate for your chosen normal and AF recording examples;
   (a) Identify the dominant frequency and any other prominent peaks in the PSD; comment on their amplitude, frequency and your understanding with respect to ECG.
2. Plot Welch's power spectral density estimate for your chosen normal and AF recording examples;
   (a) Briefly comment on the differences between both PSD methods. Why might Welch's PSD estimate be preferable for analysing biomedical signals?

4.3.1 Hints:

- To compute the modified periodogram power spectral density (PSD) estimate:

```
%% To return the modified periodogram power spectral density (PSD)
    estimate, pxx, of the input signal, x, using a hamming window:
x; %%[user set variable]: input ECG signal, measured in mV
%Note: ensure x is detrended for prior to input into the
% periodogram / FFT;
```

```
f_low; %[user set variable]: the low-pass frequency used during
    filtering previously
window=hamming(length(x)); % a Hamming window function;
%(1) modified-periodogram:
[pxx,f] = periodogram(x,window,[],fs,'psd');
%(2) Welch's power spectral density estimate
%(pwelch.m automatically implements a Hamming window):
%[pxx_pwelch,f_pwelch] = pwelch(x,[],[],[], fs,'psd');

%to plot the PSD
figure
plot(f, pxx);
xlim([0, f_low])
xlabel('Frequency (Hz)')
ylabel('Power Spectral Density (mV^{2}Hz^{-1})')
%hint: for decibels (dB) plot 10*log10(pxx) instead;
```

4.4 Exercise 4: Determining Heart Rate and Heart Rate Variability

Digital ECG signals can be analysed for events of interest which can be used to determine and monitor a patient's health status. Beyond the morphology of ECG signals, such as the shape or magnitude of the QRS complex describing the depolarisation and re-polarisation of the heart chambers, ECG signals are frequently non-stationary meaning that their frequency content changes over time. Therefore the timing of these events plays a fundamental role in understanding or monitoring how various diseases can affect the heart. Heart rate, or the number of heart beats per minute, varies due to the physical needs of the body. The rhythm of the heart is controlled by the sinoatrial nerve which is regulated by sympathetic and parasympathetic activity, the two main division of the autonomic nervous system (ANS).

Calculating the heart rate is the most basic time-series based measurement we can collect. There are many methods and complex algorithms to determine heart rate in the literature, however a threshold-based detection of the large R-peak is the simplest approach. This section will make use of the existing MATLAB `findpeaks.m` function. See MATLAB documentation for further details.

I: Determining Heart Rate through R-Peak Detection

Using the same standardised, detrended and band-pass filtered data from normal and AF recording examples in exercise A.2.3:

1. Perform R-peak detection using the MATLAB `findpeaks.m` function for your chosen normal and AF recording examples;
   (a) Choose suitable parameter values for the `findpeaks.m` function in order to best identify normal and AF R-peaks correctly;
   (b) Justify your parameter value choices;
   (c) Plot the normal and AF recording examples with the with the automatically-detected R-peak locations annotated;
   (d) Discuss the limitations of this approach to R-peak detection. Briefly comment on any other methodologies for R-peak detection you have found in the literature, including a citation from a relevant paper;

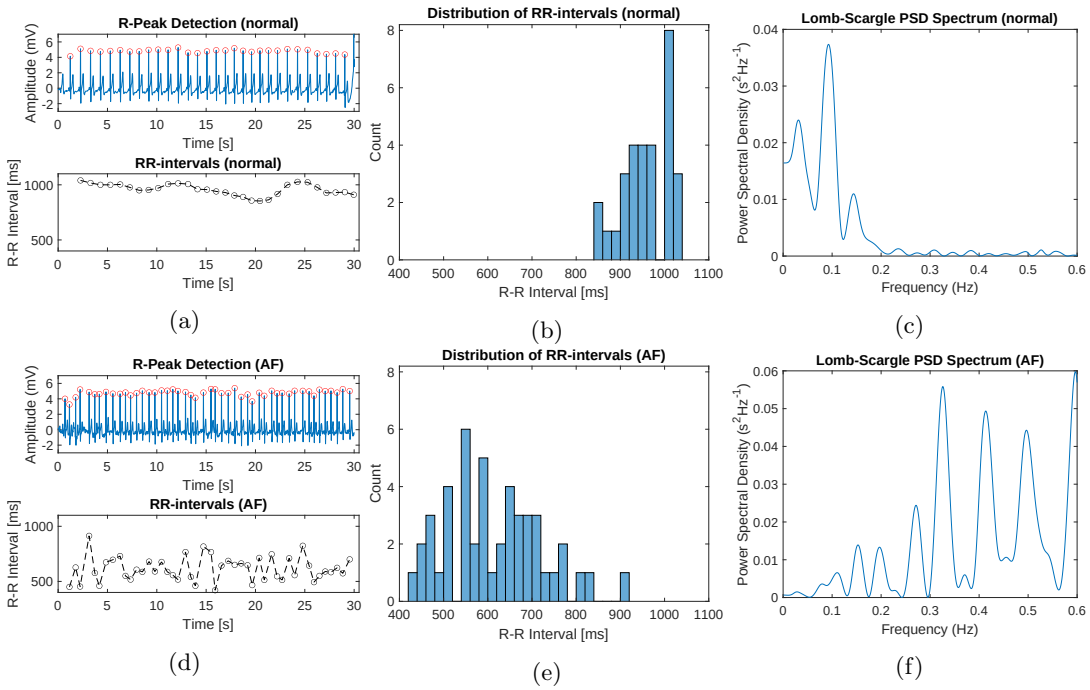


Figure 9: Panel plot demonstrating the calculation of heart rate and heart rate variability (HRV) measures from ECG. The top row (a–c) depicts a normal ECG, with an average heart rate: 62.83 beats/min; the bottom row (d–f) depicts a AF ECG, with an average heart rate: 100.80 beats/min. Panels (a) and (d) represent an annotation of the automatically detected R-peaks, with the corresponding RR-interval [ms] between successive beats for (a) normal and (d) AF rhythms respectively; panels (b) and (e) represent the distribution of the RR-interval series as histograms for (b) normal and (e) AF rhythms; the Lomb-Scargle periodogram PSD estimation of RR-intervals are represented for (c) normal and (f) AF rhythms.

2. Determine the mean heart rate from the automatically detected R-peaks for your chosen normal and AF recording examples;

II: Determining Heart Rate Variability Measures

Heart rate variability (HRV) analysis aims to understand the variation in time between each heartbeat [9]. Cardiac autonomic regulation modulates the heart rate, the variation of heart rate is mainly mediated by the sinus (sino-atrial) node, acting on heart-brain communications as well as the autonomic nervous system (ANS). HRV reflects regulation of a range of subtle physiological processes for example autonomic balance, blood pressure (BP), gas exchange, or vascular tone, which refers to the diameter of the blood vessels that regulate BP.

To evaluate HRV the variation in the R-peak-to-R-peak interval (RR) sequence of the ECG needs to be quantified. Traditional HRV analysis can be separated into two time domain measures and frequency domain measures [9, 10]. In this example you will quantify common HRV measures from the (1) time-domain: the distribution, mean and variance in heart rate; and (2) the frequency domain: estimates of the distribution of absolute or relative power into certain frequency bands. Heart rate oscillations can be divided into: very low frequency (VLF) band (0.0033–0.04 Hz), low frequency (LF) band (0.04–0.15 Hz), and high frequency (HF) components (0.15–0.40 Hz). It hypothesised that during "at rest" (i.e. normal sinus RR-intervals), high frequency variations in sinus rhythm (RR-intervals)

reflect parasympathetic (vagal) modulation of respiration (0.15 - 0.40 Hz). Higher frequencies of > 0.5 Hz typically reflect non-normal sinus rhythms (i.e. not at rest), such as during strenuous exercise. Low frequency variations in RR-intervals (0.04 - 0.15 Hz) are physiologic oscillations associated with a combination of both parasympathetic and sympathetic modulation and non-autonomic factors.

As we learned in section 4.3, a periodogram can be used to determine the frequency representation of an ECG. The RR interval sequences however will likely be unevenly sampled (i.e. the RR intervals are not uniformly spaced as they are not occurring at the same rate). The Lomb-Scargle periodogram instead has been shown to produce a more accurate estimation of the PSD than DTFT-based periodograms for unevenly sampled data [11]. For a vector $\mathbf{x} \in \mathbb{R}^{1\times N}$ (the RR intervals) with x_n, sampled at times t_n, where $n = 1, \ldots, N$, the Lomb-Scargle periodogram PSD is defined as:

$$\hat{P}_{ls}(f) = \frac{1}{2\sigma^2}\left\{\frac{\left[\sum_{n=1}^{N}(x_n - \overline{\mathbf{x}})\cos(2\pi f(t_n - \tau))\right]^2}{\sum_{n=1}^{N}\cos^2(2\pi f(t_n - \tau))} + \frac{\left[\sum_{n=1}^{N}(x_n - \overline{\mathbf{x}})\sin(2\pi f(t_n - \tau))\right]^2}{\sum_{n=1}^{N}\sin^2(2\pi f(t_n - \tau))}\right\} \tag{4}$$

where $\overline{\mathbf{x}}$ is the mean of $\mathbf{x}$; σ^2 is unbiased estimate of the variance in $\mathbf{x}$; the time offset. The time delay τ is determined by:

$$\tan(2(2\pi f)\tau) = \frac{\sum_{n=1}^{N}\sin(2(2\pi f)t_n)}{\sum_{n=1}^{N}\cos(2(2\pi f)t_n)} \tag{5}$$

Any shift $t_n \to t_n + T$ in the time measurements results in an identical shift in the offset: $\tau \to \tau + T$

1. Plot the R-R interval times over the duration of the recording from the automatically detected R-peaks for your chosen normal and AF recordings; See figure 9 for an example of this;
2. Plot a histogram of the R-R interval times over the duration of the recording from the automatically detected R-peaks for your chosen normal and AF recording examples;
3. Plot the Lomb-Scargle periodogram PSD of RR-intervals for your chosen normal and AF recording examples;
4. Briefly discuss your findings comparing your chosen normal and AF recording examples in exercise A.4.1 and A.4.2 with respect to heart rate, R-R interval times and heart-rate variability. Compare these findings to what you might expect to observe.

4.4.1 Hints:

- If your chosen ECG recording has artefacts it is often pertinent to return to exercise A.2.a and redetermine a filtering setting with a lower low-pass threshold.
- It is recommended to use the `'MinPeakHeight'` and `'MinPeakDistance'` parameters in `findpeaks.m` to initially threshold R-peaks. Further usage of this function can be found in the MATLAB documentation. Remember, the threshold of `'MinPeakDistance'` will be represented in samples, not seconds; as such you may need to convert your threshold from seconds to number of samples, again using $f = 1/t$. Experiment with other parameter settings of `findpeaks.m` as you wish;

```
% findpeaks usage
signal; %[user set variable]: the input ECG signal;
time;   %[user set variable]: monotonic vector of time values;
%[user set variable]: Your chosen scalar threshold values:
MinPeakHeight; MinPeakDistance;
%findpeaks.m function will output the R-peaks detected (R_pks)
%and the locations of the peaks (R_locs);
```

```
    [R_pks,R_locs]=findpeaks(signal, 'MinPeakHeight',MinPeakHeight,
        'MinPeakDistance', MinPeakDistance);

    %plot the R-peak annotations;
    figure
    plot(time, signal); hold on;
    plot(time(R_locs), signal(R_locs), 'ro');
```

- Below is the usage for constructing a simple histogram; See MATLAB documentation for further details.

```
    %to get the RR-intervals in seconds use the diff.m function
     %diff(X) calculates differences between adjacent elements of X
     %Calculate the difference between consecutive RR intervals;
     %convert samples to seconds by dividing by the sampling rate;
     RRs=diff(R_locs)./fs;
     %convert from [s] to [ms]:
     RRms=RRs.*1000

     %% Constructing a simple probabilistic histogram
     BinWidth=20; %[user set variable]: the width of the bins
     figure
     histogram(RRms, 'BinWidth', BinWidth)
     xlabel('R-R Interval [ms]')
     ylabel('Count')
```

To evaluate the Lomb-Scargle periodogram PSD of RR-intervals using the plomb.m MATLAB function. See MATLAB documentation for further details.

```
    %Lomb-Scargle PSD RR-interval Spectrum;
    RRs; %The RR-interval in seconds [s];
    RRs=RRs-mean(RRs); %detrend the RR intervals (best practice);
    %Determine the time points in the ecg each R-peak occurred
    %from the first interval (i.e. the second R-peak detected);
    t=time(R_locs(2:end));
    %Denote the frequencies we are interest in evaluating:
    %[0.001, 0.6] Hz;
    f_interest=0.001:0.001:0.6;
    % Compute the Lomb-Scargle periodogram;
    [pxx_plomb,f_plomb] = plomb(RRs, t, f_interest, 'psd');

    %to plot the PSD
    figure
    plot(f_plomb, pxx_plomb);
    xlabel('Frequency (Hz)')
    ylabel('Power Spectral Density (s^{2}Hz^{-1})')
```

- For you information: simple HRV time-domain and frequency-domain measures can be extracted for recording examples using the provided MATLAB function `hrv_measures.m`; usage:

```
    %RRms=RRs*1000: A vector of RR intervals, converted from seconds
        to miliseconds [ms];
```

```
    %f_plomb:  The frequency vector returned by the Lomb-Scargle
    %periodogram;
    %pxx_plomb: The Lomb-Scargle power spectral density (PSD)
    %estimate at f;

    %Function to return Heart Rate Variability (HRV) measures:
    HRV=hrv_measures(RRms, f_plomb, pxx_plomb);
    %display the HRV measures:
    display(HRV)
```

5 Section B: The Electroencephalogram

In this section you will learn some standard approaches for analysing ectroencephalogram (EEG) data. The most common method for EEG scalp placement is the 10-20 system. The 10 and 20 refer to the percent distances between the electrodes in proportion to the size of the head. Figure 10 depicting EEG electrode placement following the 10-20 convention and corresponding EEG trace taken from an infant during sleep and recorded under clinical supervision in an Neonatal Intensive Care Unit (NICU). EEG (μV) signals are much smaller than ECG (mV) and therefore data collection and pre-processing steps play an pivotal role in EEG analysis. In the next section you will learn how to correctly process experimentally obtained EEG data.

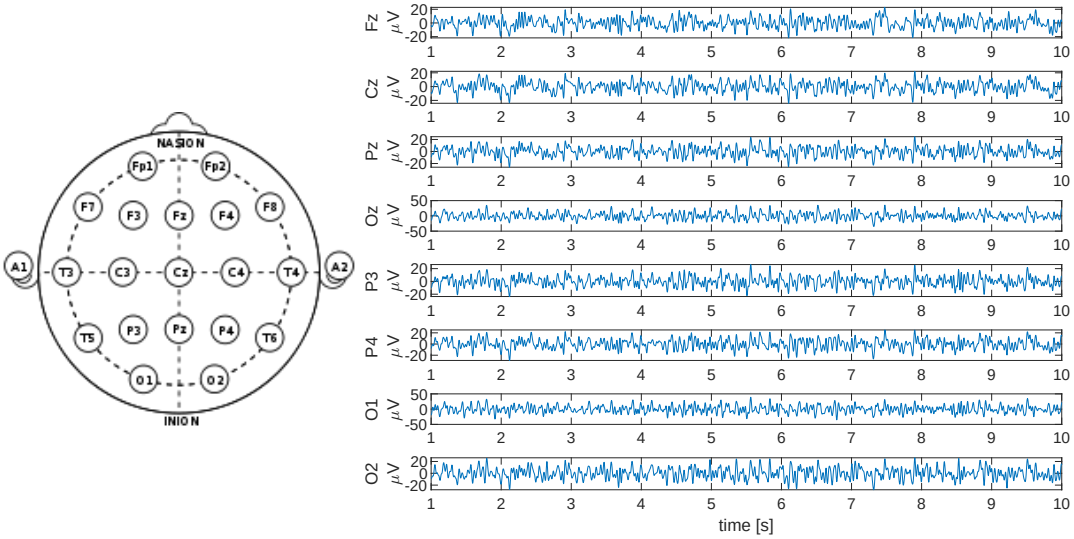


Figure 10: Example depicting EEG electrode placement following the 10-20 convention and corresponding EEG trace taken from an infant during sleep.

5.1 Exercise 1: Investigating EEG Data

This example exercise uses example data from a healthy participant (age 24 years) with electrodes placed across the occipital area of the brain (O1-Cz), sampled at 256 hertz (Hz). In the experiment, EEG from the participant was acquired with the subject at rest with their eyes open (trial #1). Next, the participant was asked to close their eyes as EEG data was again collected (trial #2). It is expected that students will utilise their MATLAB leanings from section A to complete this section.

1. Load the EEG data: `eeg_trial_1.mat`; `eeg_trial_2.mat` and plot the entire EEG samples for each for the first 512 samples of each trial;
2. Pre-process the data by low-pass filtering, detrending, and normalising each EEG signal (apply these steps in series).
   (a) Determine suitable low-pass filtering parameters by reporting and analysing various parameter combinations;
   (b) Justify your filtering parameters choices with respect to your understanding of analysing EEG signals;
   (c) Plot the EEG samples for the first 512 samples of each trial (#1 & #2) after pre-processing;

3. Plot the modified periodogram PSD estimate of each trial (display only up to 60 Hz);
   (a) At what frequencies are the four highest components remaining after filtering? Suggest a cause for each;
   (b) Note, discuss, and compare any significant peaks in each spectrum comparing trial #2 and trial #2;
4. Describe the main sources of error in EEG experiments. Suggest methods to mitigate or control for these identified sources of error.

5.1.1 Hints:

- Design a Butterworth Low-Pass Filter:

```
%Design a LPF
fs; %[set this value]: the sampling frequency in Hertz (Hz)
fc;  %[set this value]: cut-off frequency in Hertz (Hz)
norder; %[set this value]: filter order
%The cutoff frequency Wn must be 0.0 < Wn < 1.0;
Wn=fc/fs;
%Design the filter
[B,A] = butter(norder, Wn);
%apply the LPF
signal=filter(B,A,signal);
```

6 Section C: Smartphone Physical Activity Monitoring

In this section you will investigate smartphone accelerometry from the annotated accelerometer ubicomp2013 dataset by Brajdic and Harle (2013) [12]. Further information on the dataset can be found at: `http://www.cl.cam.ac.uk/~ab818/ubicomp2013.html`. The ubicomp2013 study collected smartphone inertial measurement unit (IMU) data (tri-axial accelerometer, gyroscope, magnetometer sensors) from a number of participants as they walked. The smartphone was placed in different locations on the body across different experiments. These included the front and back pockets, held in the subject's hand or in a backpack. In this lab we are only interested in smartphone 3-axis accelerometer data and those experiments performed with the smartphone in the front pocket. Figure 11a depicts an example of a smartphone 3-axis accelerometer trace captured from a participant in the ubicomp2013 study with the smartphone in the front pocket. The ubicomp2013 study was performed under video observations, so importantly each experiment file has been annotated for the actual number of steps taken and the times when the walk began and ended.

The simplest gait-based health indicators are daily-step counts, which estimate the number of steps per-day and therefore the daily activity and activity intensity. To build a step detection algorithm we must first identify periods of walking (termed a bout). Next we perform step counting on those detected bouts. Figure 11 demonstrates the walking detection and step counting pipeline.

You are provided with two MATLAB files for exercise 1 and 2: `B18_GAIT_SectionC_Exercise_1.m` and `B18_GAIT_SectionC_Exercise_2.m` respectively, which will allow you to perform the analysis for this section. These files are pre-programmed to load and analyse the data files, however you will have to change some of the parameter options.

Walking Detection Algorithm

In this lab we will use a heuristic[1] walking detection algorithm based on the approaches introduced in [13, 14]. This method utilises a threshold approach to detect period of walking:

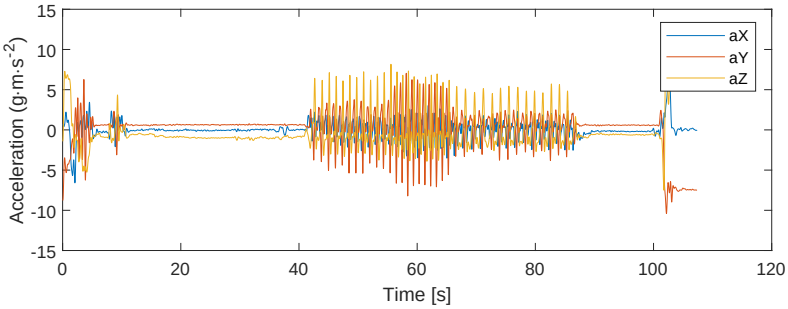
1. First a moving average window of the magnitude of acceleration (aM) is determined;
2. The moving average aM is compared to a threshold value `acc_threshold` to detect periods of movement;
3. A moving average window of the summed standard deviation (ssd) across tri-axial accelerometer axis is calculated (i.e. the sum of the standard deviation of each $\boldsymbol{a}_x$, $\boldsymbol{a}_y$, and $\boldsymbol{a}_z$);
4. The moving average ssd is compared to a threshold value `ssd_threshold` to detect periods of movement variation;
5. Periods of walking are determined by $aM >$`acc_threshold` and $ssd >$`ssd_threshold` simultaneously.

Figure 11b shows an example of smartphone-based walking detection, with figure 11c demonstrating the parameterisation of the walking detection algorithm.

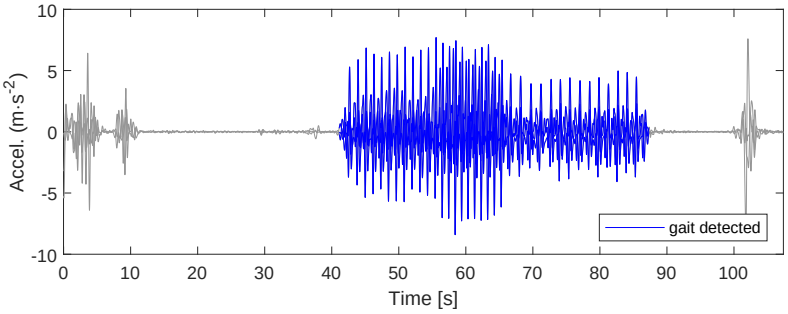
Step Counting Algorithm

There are many methodological approaches to detect steps from wearable accelerometry. The simplest is a heuristic peak detection, much like you implemented during section A exercise 4, which can also be evaluated using the `findpeaks.m` function in MATLAB. You will need to determine the threshold

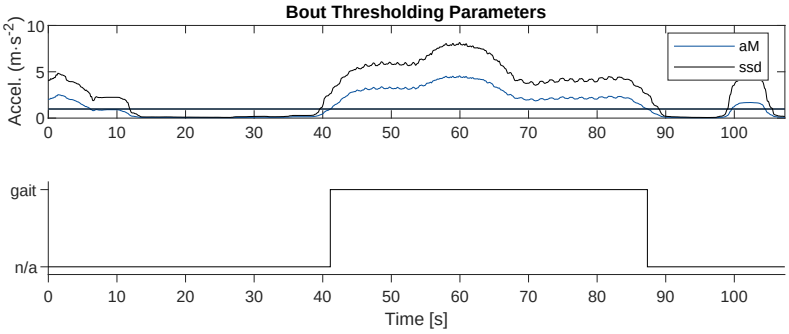
[1] From Wikipedia: "A heuristic technique, is any approach to problem solving or self-discovery that employs a practical method that is not guaranteed to be optimal, perfect, or rational, but is nevertheless sufficient for reaching an immediate, short-term goal or approximation."



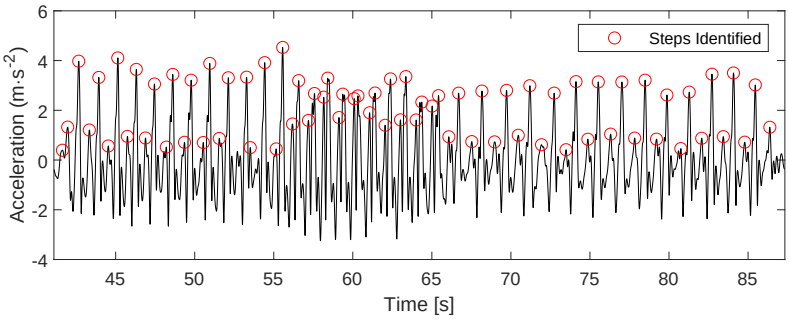
(a) Smartphone 3-axis accelerometer data



(b) Walking Detection



(c) Walking Detection Algorithm Parameters



(d) Step Detection

Figure 11: Smartphone walking identification and step detection algorithm pipeline. Figure (a) depicts the raw smartphone 3-axis accelerometer data; (b) highlights the region of (processed) sensor signal identified as walking; (c) demonstrates the parameterisation of the heuristic walking detection algorithm: the moving average acceleration (aM) and the combined standard deviation (ssd) threshold; (d) annotates the steps automatically identified through peak detection across the detected walking bout.

values for step detection: `MinPeakDistance`, the minimum distance between consecutive peaks, in seconds [s]; the `MinPeakHeight` the minimum height of a step peak, measured in $[m \cdot s^2]$. Figure 11d demonstrates heuristic peak identification for step detection from an identified walking bout.

6.1 Exercise 1: Building a Smartphone Walk Detection and Step Counting Algorithm

The goal of this exercise is to understand how the parameters for walking detection and step counting can be determined heuristically. You are provided with the MATLAB file `B18_GAIT_SectionC_Exercise_1.m`. You should aim to compare parameter choices across multiple different files:

1. Run the `B18_GAIT_SectionC_Exercise_1.m` file for any example file of your choice;
   (a) Pick suitable `acc_threshold` and `ssd_threshold` values for walking detection;
   (b) Pick suitable `MinPeakDistance` and `MinPeakHeight` values for step detection;
   (c) Record your parameter values;
   (d) Record the actual number of steps, the number of steps detected and the cadence for those parameter values;
2. Similar to figure 11, for your chosen file, plot:
   (a) raw 3-axis smartphone acceleration signal;
   (b) the walking detection outcome;
   (c) the walking detection algorithm parameters chosen;
   (d) the step detection annotations
3. Repeat steps 1 and 2 for at least one other recording;

6.1.1 Hints:

- To load the data, you will need to change the '`pathname`' variable to the directory on your computer where the data is stored:

```
%% Load the Data
%the pathname where the data is stored; You will need to change this to
%your directory where the data is stored, e.g.:
pathname='C:\MATLAB\DATA\';

%set the device location; in this example we are only interested in
%smartphone accelerometry that recorded when the phone was in the front
%trousers pocket.
device_location='front_pocket';
% returns a table consisting of the filenames extracted and the
% corresponding subject ID as strings.
fileInfo=return_filenames_gait(pathname, device_location);
```

- See below an example for how to change the parameter values:

```
G=9.81; %[fixed variable]: gravity 9.81 m/s^2
options.acc_threshold=0.15*G; %[user set variable]: set the moving
    average accel. threshold value to 15% of gravity
options.ssd_threshold=0.3*G;  %[user set variable]: set the moving
    average accel. threshold value to 30% of gravity
```

```
options.MinPeakDistance=0.5;  %[user set variable]: set to 0.5 [s]
options.MinPeakHeight=0.1*G   %[user set variable]: set to 10% of
    gravity [m/s^2]
```

- To chose a specific file, you will also need to change the '`file_index`' variable to a value of your choice corresponding to the filename you wish to analyse;

```
%% Run Walking Detection and Step Detection Algorithm
%choose one of the filenames from the dataset and corresponding
    subject
%identifier
file_index=10; %[user set variable]: choose a value [1, 2, ..., N],
    where N are the number of files
%get the corresponding filename and subject name from the fileInfo
    table:
filename=fileInfo.filename(file_index); subject=fileInfo.subject(
    file_index);
```

6.2 Exercise 2: Assessing Step Counting Algorithm Performance

In this exercise you will assess the generalised step count algorithm performance across all participants. You are provided with the MATLAB file `B18_GAIT_Exercise_2.m`. This file will run the walking detection and step detection for all participants in the entire ubicomp2013 study (as with exercise 1, we are only interested in front pocket smartphone locations for this lab). The `B18_GAIT_Exercise_2.m` file will then plot the ubicomp2013 population distribution of the actual number steps observed versus the number of steps detected through the heuristic step detection algorithm (see figure 12). Further performance summary statistics which measure the error of our step detection algorithms are also evaluated, such as the mean absolute error (MAE) and mean squared error (MSE) between actual and detected number of steps per subject:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{6}$$

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{7}$$

where y are the steps observed; $\hat{y}$ are the steps detected and N are the number of subjects. In this example the average and standard deviation in the number of steps per subject was observed as: 77.3 $\pm$ 7.8 steps, compared 76.9 $\pm$ 13.4 steps detected by our heuristic step detection algorithm. The mean absolute error (MAE) between steps observed and steps detected was determined as 5.8 steps; the MSE between steps observed and steps detected was determined as 80.9. This is relatively good step detection performance!

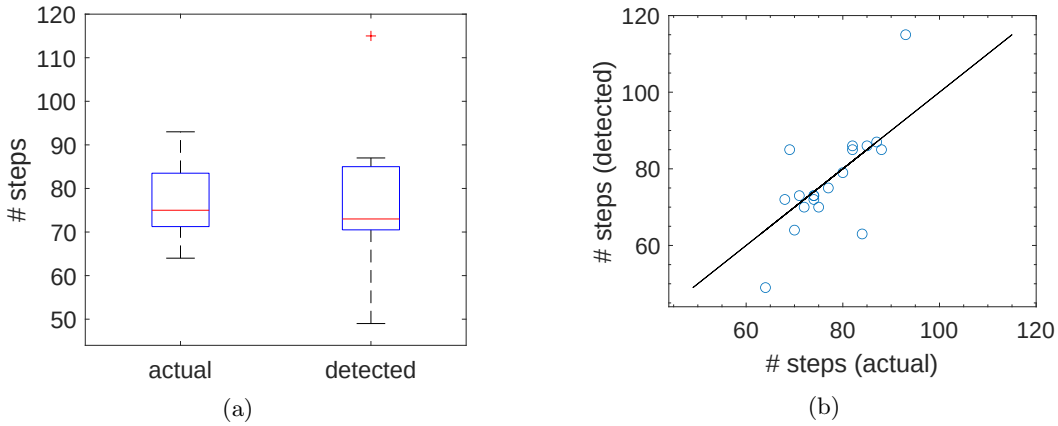


Figure 12: Analysis of step detection performance. across the ubicomp2013 study. (a) illustrates the boxplot distribution of actual vs. detected number of steps per subject; (b) depicts a scatter plot of the actual vs. detected number of steps per subject. A black line represents prefect step detection. (# steps observed: 77.3 ± 7.8 steps, # steps detected: 76.9 ± 13.4; p= 0.86; MAE: 5.8, MSE: 80.9)

1. Run `B18_GAIT_SectionC_Exercise_2.m` with your chosen parameter values for `acc_threshold` `ssd_threshold` values for walking detection and `MinPeakDistance`, `MinPeakHeight` values for step detection;
   (a) Plot the boxplot distribution of actual vs. detected number of steps per subject;
   (b) Plot the scatter plot of the actual vs. detected number of steps per subject;
   (c) Record the MAE and MSE error metrics between the actual vs. detected number of steps;
   (d) Briefly discuss the performance of your chosen parameters;
2. Briefly discuss some of the difficulties and limitations with the heuristic walking detection and step counting algorithm analysed in section C;
3. From the literature, suggest any another method for either walking detection or step counting;

6.2.1 Hints:

- Remember to modify the file `B18_GAIT_SectionC_Exercise_2.m` to include your chosen parameter values for `acc_threshold`, `ssd_threshold` values for walking detection and `MinPeakDistance`. `MinPeakHeight` values for step detection; To do this, refer to the hints for section C, exercise 1 previously.
- Don't get caught up spending time on parameter optimisation. The purpose of this example is not to obtain perfect performance, but to understand, implement, and critique simple heuristic methods that can perform walking and step detection.

References

[1] T. C. G. Fink, "Using phase response curves to understand neuronal synchronization and sleep." Ph.D. dissertation, 2012.

[2] C.-C. Yang and Y.-L. Hsu, "A review of accelerometry-based wearable motion detectors for physical activity monitoring," *Sensors*, vol. 10, no. 8, pp. 7772–7788, 2010.

[3] J. J. Kavanagh and H. B. Menz, "Accelerometry: a technique for quantifying movement patterns during walking," *Gait & posture*, vol. 28, no. 1, pp. 1–15, 2008.

[4] A. Doherty, D. Jackson, N. Hammerla, T. Plötz, P. Olivier, M. H. Granat, T. White, V. T. Van Hees, M. I. Trenell, C. G. Owen *et al.*, "Large scale population assessment of physical activity using wrist worn accelerometers: The uk biobank study," *PloS one*, vol. 12, no. 2, p. e0169649, 2017.

[5] P. F. Saint-Maurice, R. P. Troiano, D. R. Bassett, B. I. Graubard, S. A. Carlson, E. J. Shiroma, J. E. Fulton, and C. E. Matthews, "Association of daily step count and step intensity with mortality among us adults," *Jama*, vol. 323, no. 12, pp. 1151–1160, 2020.

[6] A. P. Creagh, C. Simillion, A. Bourke, A. Scotland, F. Lipsmeier, C. Bernasconi, J. van Beek, M. Baker, C. Gossens, M. Lindemann *et al.*, "Smartphone-and smartwatch-based remote characterisation of ambulation in multiple sclerosis during the two-minute walk test," *IEEE Journal of Biomedical and Health Informatics*, 2020.

[7] A. P. Creagh, "The development of digital biomarkers for multiple sclerosis from remote smartphone- and smartwatch-based assessments," Ph.D. dissertation, University of Oxford, 2020.

[8] G. D. Clifford, C. Liu, B. Moody, H. L. Li-wei, I. Silva, Q. Li, A. Johnson, and R. G. Mark, "Af classification from a short single lead ecg recording: the physionet/computing in cardiology challenge 2017," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.

[9] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Medical and biological engineering and computing*, vol. 44, no. 12, pp. 1031–1051, 2006.

[10] F. Shaffer and J. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, vol. 5, p. 258, 2017.

[11] N. R. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophysics and space science*, vol. 39, no. 2, pp. 447–462, 1976.

[12] A. Brajdic and R. Harle, "Walk detection and step counting on unconstrained smartphones," in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, 2013, pp. 225–234.

[13] A. Hickey, S. Del Din, L. Rochester, and A. Godfrey, "Detecting free-living steps and walking bouts: validating an algorithm for macro gait analysis," *Physiological measurement*, vol. 38, no. 1, p. N1, 2016.

[14] G. Lyons, K. Culhane, D. Hilton, P. Grace, and D. Lyons, "A description of an accelerometer-based mobility monitoring technique," *Medical engineering & physics*, vol. 27, no. 6, pp. 497–504, 2005.