1 Project Goals

1.1 Goals

The main goal of this project is to develop a prototype system that uses an individual's ECG (Electrocardiogram) for authentication. The system will compare the user's heartbeat with a previously registered heart signature and make a decision to accept or deny the user. The project can be broken down in to two major development phases. Phase 1 being the development of a system that can acquire and store ECG signals; phase 2 being the implementation of an ECG identification algorithm used to authenticate users by comparing the signal to a database of registered users.

An Arduino DUE microcontroller will be used as the main hardware tool for deployment, dealing with ECG acquisition and pre-processing before sending the data to a PC with Matlab installed. The ECGs will be acquired using the OLIMEX EKG/EMG shield and passive wrist electrodes, providing a single lead ECG signal. Matlab will be used to process the signals (filtering and adjustment) and store them on the PC. The filtered signals will be used by the ECG identification algorithm developed in Matlab to authenticate the user. If time permits it, improvements could be made to the final implementation by developing a more portable prototype: Using a Raspberry Pi running Python scripts as opposed to a PC running Matlab. It is unlikely that enough ECG signals will be acquired to robustly test the ECG ID algorithm, hence a database of pre-recorded ECGs will be used (provided by physionet.org) for this purpose. The pre-recorded lead I signals can be imported directly into Matlab.

1.2 Justification for Approach

An Arduino platform was chosen due to compatibility with the OLIMEX ECG shield. The shield was selected after extensive research and evidence that others had success with it in past projects. The OLIMEX shield has the added benefit of supporting the passive wrist electrodes. This allows ECG data to be acquired in a non-invasive manner, an important factor if useful applications of this technique are to be viable. The Arduino DUE was selected over the more common Arduino UNO as it has a more powerful processor (GIVE NUMBER COMPARISON) and a higher resolution ADC: 12-bit compared to 10-bit The pre-recorded ECGs also have a 12-bit resolution making the acquired ECGs consistent with the test data for the ID algorithm.

Matlab was chosen due to its powerful signal processing capabilities and community support for filtering. Interfacing Matlab with an Arduino connected to a serial port is fairly trivial, providing another strong argument for the use of Matlab. Programming with embedded C and Matlab corroborates with my skill set.

1.3 Scope

This project intends to add to existing research in a novel approach to authentication, through exploring various techniques used for ECG acquisition and identification. To this end, a specific application will not be implemented, however if there is enough time, the prototype could be applied to a useful situation for demonstration purposes. For example, a payment or login system.

2 Background and Report of Literature Search

2.1 Problems with Current Authentication Techniques

The most widely used form of authentication, the password, has a large number of problems suggesting its dominance should be coming to an end. Herley et al. [?] explores these downfalls citing several weaknesses in passwords including their vulnerability to brute force, phishing, man-in-the-middle, social engineering and key-logging attacks, forgetfulness, multiple use and exposure following data breaches [?]. This incentivises the adoption of biometrics as a form of authentication. Biometrics combat many of the discussed issues associated with the use of passwords. Their main advantages being their uniqueness, difficulty to replicate and the fact that they cannot be lost or forgotten. Biometrics can contain drawbacks. Some can be stolen or replicated such as fingerprints, iris and face (through use of a photograph); these commonly used biometrics also contain an inherent flaw: They have no measure of liveliness, leaving them exposed to the possibility of theft or replication [?]. The ECG, on the other hand, is fundamentally a measure of liveness, cannot be spoofed and can be used as a biometric [?].

2.2 ECG Mechanics

ECG signals measure the changes in electric potential generated by the heart. Conductive electrodes on the skin measure the electrical depolarisation of the heart. An ECG trace contains three major complexes (P, R, T). These are defined by the fiducial points of the peak of each complex [?]. figure 1 shows an example ECG trace with fiducials, commonly used by the medical community, labelled. The orientation, size and strength of the heart affect the placement of these fiducials [?]. The trace is also affected by anatomic features as the electrical currents generated by the heart spread around the body [?]. These factors give ECG traces their individuality.

2.3 ECG Acquisition

There are 12 possible virtual vectors referred to as leads used to record an ECG [?]. A lead I ECG is a lateral vector going from left to right across the chest, it only requires two electrodes, but can be used with three (an additional ground or DRL electrode). figure 2shows the lead I vector. Using the lead I vector for electrode placement permits a non-invasive approach to acquiring ECGs, as they can be taken from the wrists of the user. Biel et al [?], performed a study that first proved that it is possible to identify a person using their ECG and concluded that a lead I ECG, using three electrodes is good enough for identification purposes. The Driven Right Leg (DRL) electrode, marked in red in figure 2, is used to suppress electrical interference [?]. The DRL electrode is connected to the output of an auxiliary op-amp to form a negative feedback loop, using the common-mode voltage of the body as an input. A small amount

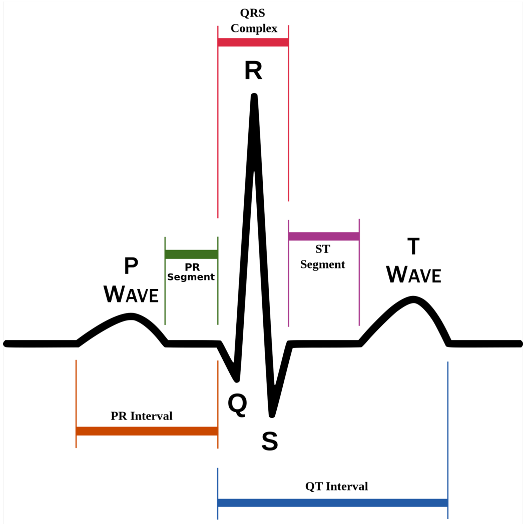

Figure 1: Ideal ECG trace with complexes labelled. [?]

of current (<1uA) is driven to the DRL electrode to match the displacement currents in the body. The feedback form the circuit reduces the common-mode voltage [?].

2.4 ECG Signal Processing

Raw ECG data is often quite noisy and can be distorted. It is therefore necessary to employ various filtering techniques to extract a clean ECG signal like the one in figure 1. Common sources of noise include: power line interference, EMG noise, noise caused by instrumentation and electrode contact noise. figure 3 provides a summary the main noise sources and their method of introduction into the signal. An ECG signal taken from a rested patient operates in the 0.67-1.7Hz range (between 40 and 100 beats per minute) [?]. When processing a raw ECG signal the goal is to isolate the frequencies in this range and filter out other frequencies.

Power line interference is caused by capacitive and inductive coupling and occurs at close to 60Hz. Capacitive coupling occurs when two circuits transfer energy through coupling capacitance between them [?]. Capacitive coupling is reduced by increasing the distance between the circuits or increasing the shielding between them. Inductive coupling is caused by the mutual inductance between conductors. The magnetic flux produced by current flowing through a wire can induce a current in nearby circuits [?]. Inductive coupling is the main source for low frequency noise, making it the dominant factor in power line interference in ECGs [?]. The effects of inductive coupling are reduced in a similar way to capacitive coupling prevention. Additional counter measures

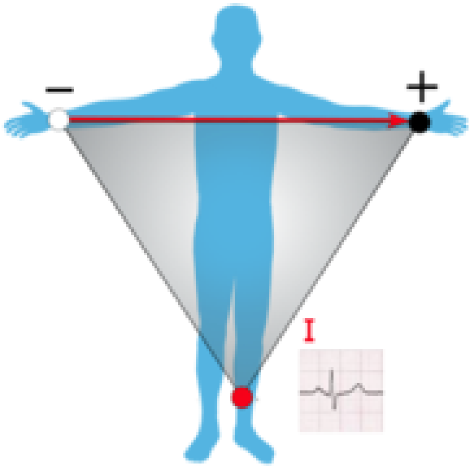

Figure 2: Lead I ECG vector [?]

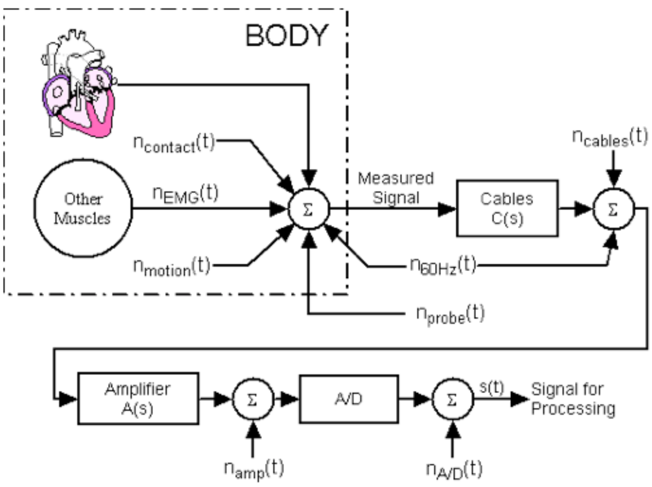

Figure 3: The principal noise source in an ECG [?].

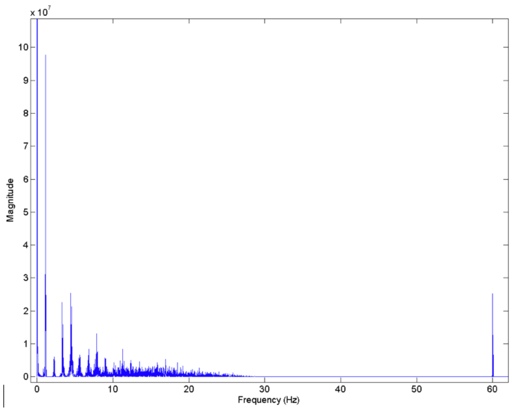

Figure 4: The Fourier power spectrum of a raw ECG trace with 60Hz power line noise visible [?].

include ensuring the electrodes have a good connection and the use of twisted wires. figure 4 shows a fourier power spectrum of a raw ECG trace, with the power line interference clearly identifiable at 60Hz. This is corroborated by the high frequency noise signal observable in Figure X, showing a raw ECG trace recorded using the setup described in section 1.1.

EMG noise is caused by muscles other than the heart contracting; causing depolarisation and repolarisation waves that can be observed in the ECG trace [?]. This can be prevented by the subject keeping as still as possible and using good quality probes.

Electrode contact noise causes low frequency (~0.06Hz) baseline drift and changes in amplitude of the ECG trace. A reduction in the amplitude of the signal increases the probability of noise by reducing the SNR (signal to noise ratio). figure 4 shows an example of the effects of baseline drift on an ECG trace. One method for correcting baseline wander is detailed by Zhang [?]. This method involves estimating the baseline wander by coarse approximation in discrete wavelet transformation. Zhang suggests using the Symlets wavelet with order 10, owing to its similarity with a QRS complex. A visual and mathematical approach are suggested for determining the decomposition level. The mathematical approach involves evaluating the frequency range for each coarse approximation decomposition level and using one that spans the largest frequency below 1Hz (typical frequency for baseline wandering). This is then subtracted from the original signal to remove the baseline wandering.

Filtering the high and low frequency noise components can be achieved by using a bandpass filter.figure 6 shows how effective a bandpass filter is in removing the unwanted noise frequencies. Lugovaya formalises the data preprocessing of an ECG in the following way: baseline drift correction, frequency-selective filtering and signal enhancement [?].

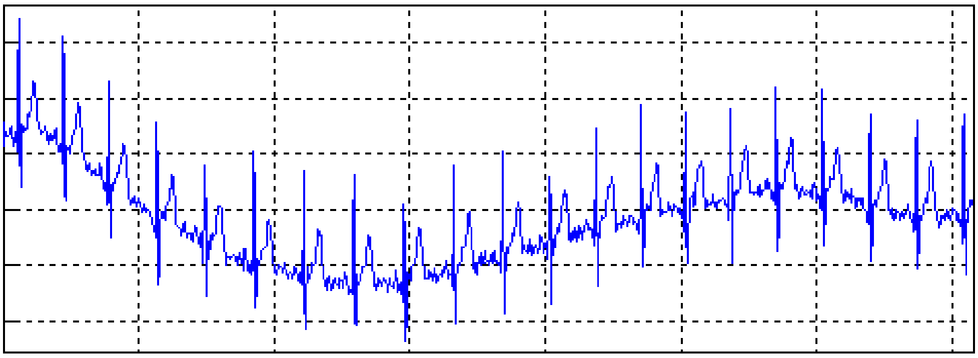

Figure 5: ECG trace with baseline drift [?]

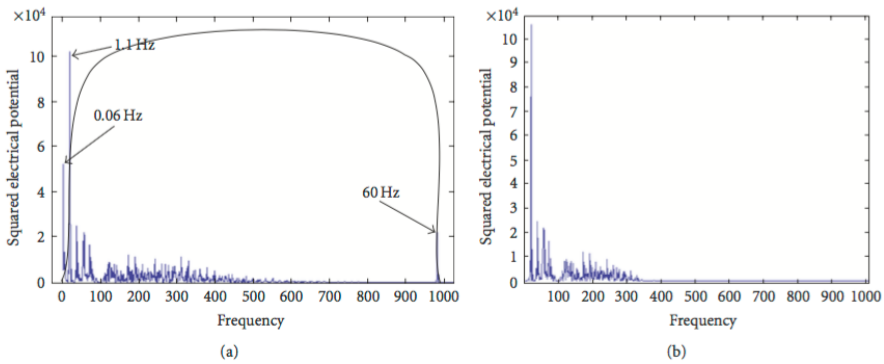

Figure 6: Power spectra of: (a) bandpass filtering raw ECG (b) filtered data[?]

2.5 ECG Identification

Several different approaches have been made in the development of an identification algorithm. All start with feature extraction and characterization of the ECG signal as the standard medical characterization of the fiducial points of an ECG is not comprehensive.

Biel et al. used the trace's shape to identify people by extracting amplitude and time based features for classification [?]. The results proved the possibility for using an ECG as a biometric.

A sequential procedure is presented by Israel et al., where fifteen time ECG attributes are extracted [?]. These features comprised of distances between various peaks and valleys in the trace. The study aimed to investigate the effect of heart rate changes on the ECG features. Linear normalization was used, where the length of each feature is divided by the total length of the heartbeat, to remove the effects of a change in heart rate. Conventional voting techniques were used for identifying subjects.

Singh and Singh proposed a more complex approach for identification [?]. Twenty features were extracted, from each beat in the ECG trace, based on: amplitude, time and angles. The Euclidean distance is then calculated between the features in the enrolled trace and the authenticating trace. A decision to accept or reject the user is made base on these values. The results showed a TAR (true acceptance rate) of 82% and FAR (false acceptance rate) of 7%.

Lugovaya uses a similar approach to Israel et al. by employing a linear discriminate analysis and majority vote classifier to identify subjects [?]. The study concludes with a TAR of 96% and FAR of 0.5%. Compared to similar studies these results are very good.

Arteaga-Falconi et al. use a method that involves extracting six time and two amplitude based features from a PQRST segment to use in an identification algorithm [?]. The algorithm compares each time-based feature from the enrollment template with the authentication template. If the features match within a certain tolerance a counter is incremented. The counter's score is then evaluated, if it is below a certain score 'no match' is returned, if it is above or below the minimum score the amplitude based attributes are evaluated. These must both be a close match to the enrollment template for 'match' to be returned. A TAR of 81.81% and FAR of 1.41% was achieved with a signal acquisition time of just four seconds.

3 Report on Technical Progress

3.1 Acquiring ECG Signals

ECG signals can be acquired using the setup discussed in section 1.1. An Arduino DUE is being used with the OLIMEX ECG shield and passive wrist electrodes to obtain an ECG trace. The Arduino was programmed using the code found in Appendix A. This code was sourced from the OLIMEX forum for the shield. The program sets up a timer that fires every 1/250 of a second. This causes the ADC value on the Arduino to be read and then sent to the serial port.

Initially, the signals were read and processed by the Arduino using the Arduino IDE serial monitor to output the signal. figure 7 shows a screenshot of the serial output monitor. The signal is noisy and has a low amplitude due to poor electrode contact. Tightening the wrist straps improved the quality of the signal. The improved signal is shown in figure 8, the R peak has a much better definition.

A serial connection was later established between Matlab and the Arduino. Appendix B shows the program that helped achieve this. The program sets up the serial connection with the Arduino, records the data it gets sent into a vector for 10 seconds. The serial communication is then closed, the data is plotted and the workspace is saved.

3.2 ECG Signal Processing

The raw ECG signals are filtered using the ECG_filter function shown in appendix C. The function uses three different signal filtration techniques available in Matlab and plots the output for each one. figure 9 shows the output for each filter along with the original data. The weighted window and gaussian filters have removed most of the high frequency noise whilst maintaining the signals integrity. The butterworth filter does not completely remove the high filter noise, giving the impression of extra fiducials. The gaussian filter preserves more of the signals amplitude than the weighted window filter.

3.3 Importing Pre-Recorded Signals

The physionet.org physiobank database [?] contains 310 lead 1 ECG recordings taken from 90 different subjects [?]. These signals can be imported into Matlab using the 'wfdb toolbox'. Figure X shows an imported ECG trace from the database. This provides easy access to ECG signals without the need for participants or reliance on hardware. The test data is ideal for testing and developing the I.D algorithm.

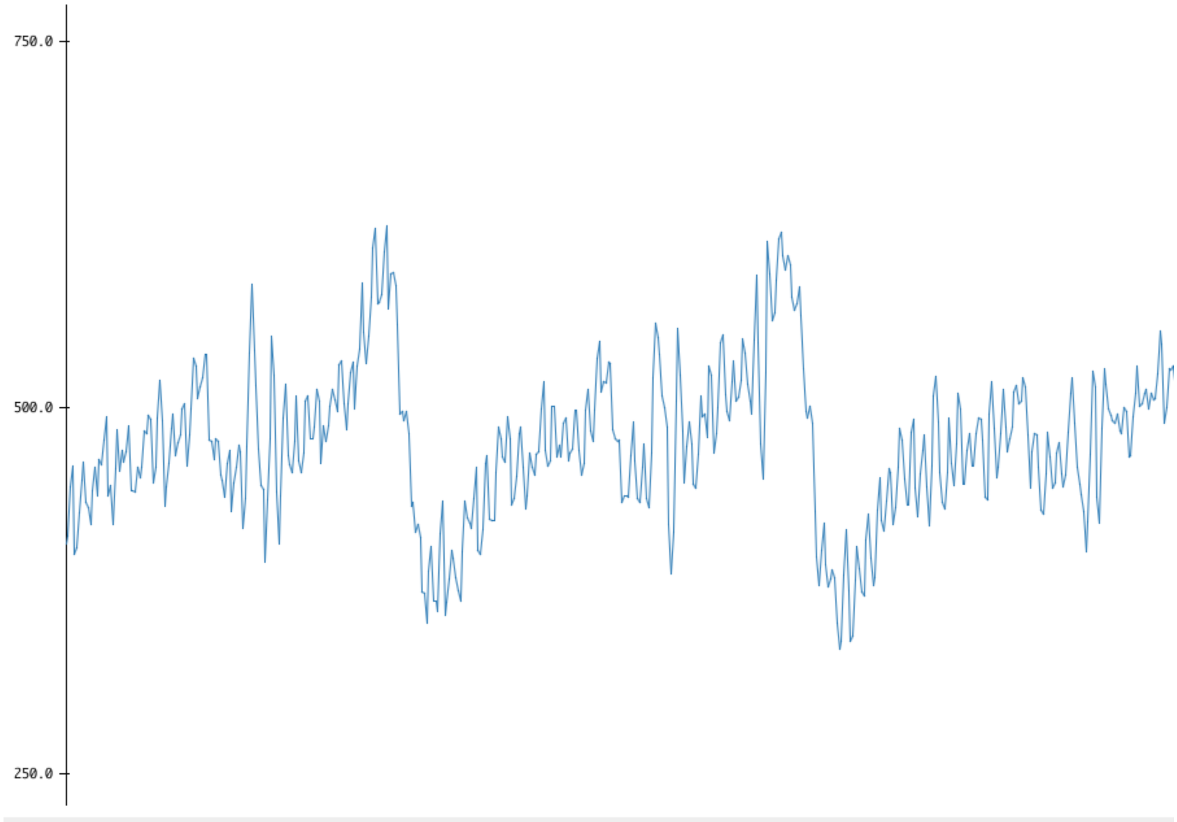

Figure 7: Noisy serial output from Aduino DUE with OLIMEX ECG shield

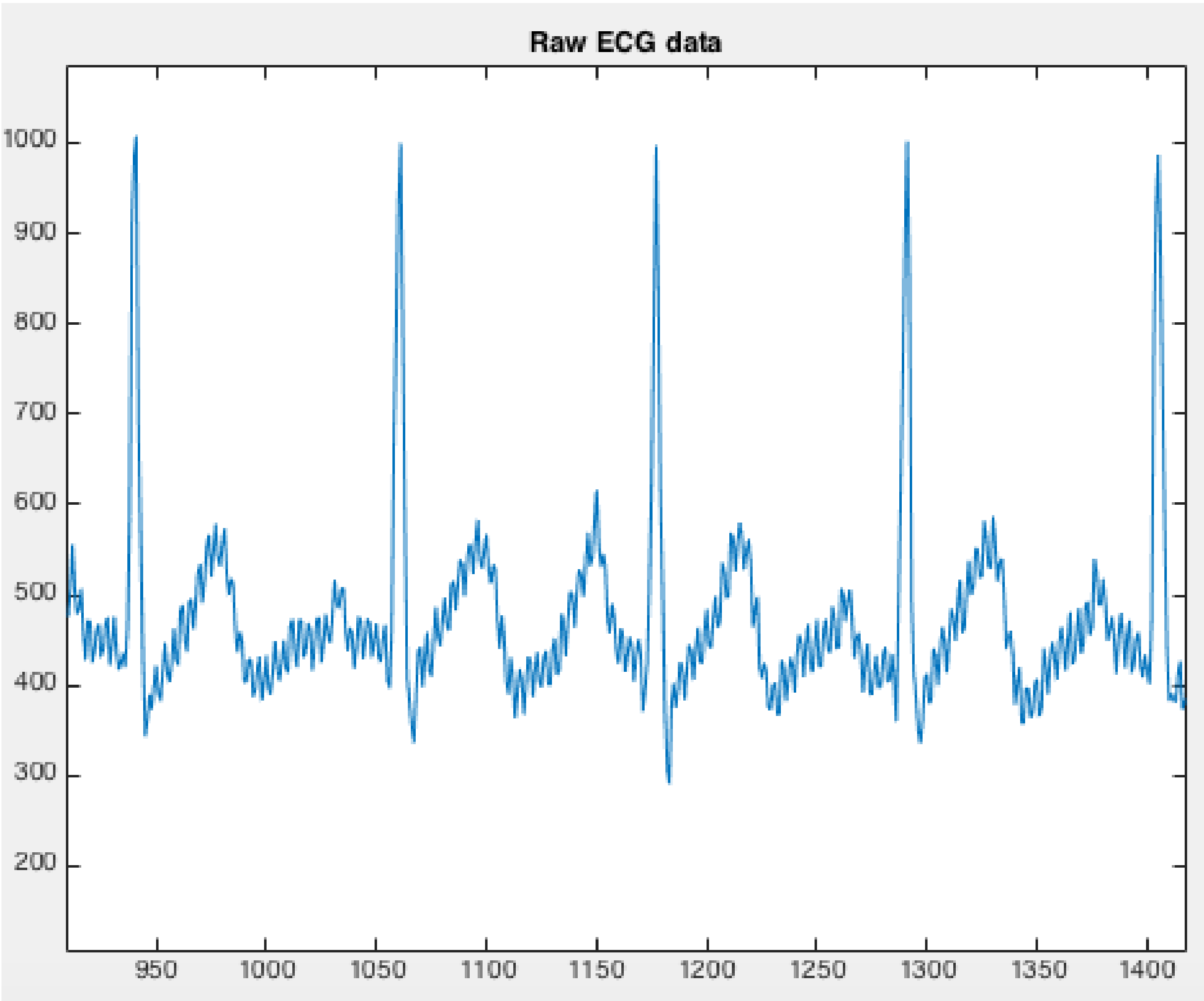

Figure 8: Improved ECG signal from Arduino Due and OLIMEX ECG shield. Output is from MATLAB

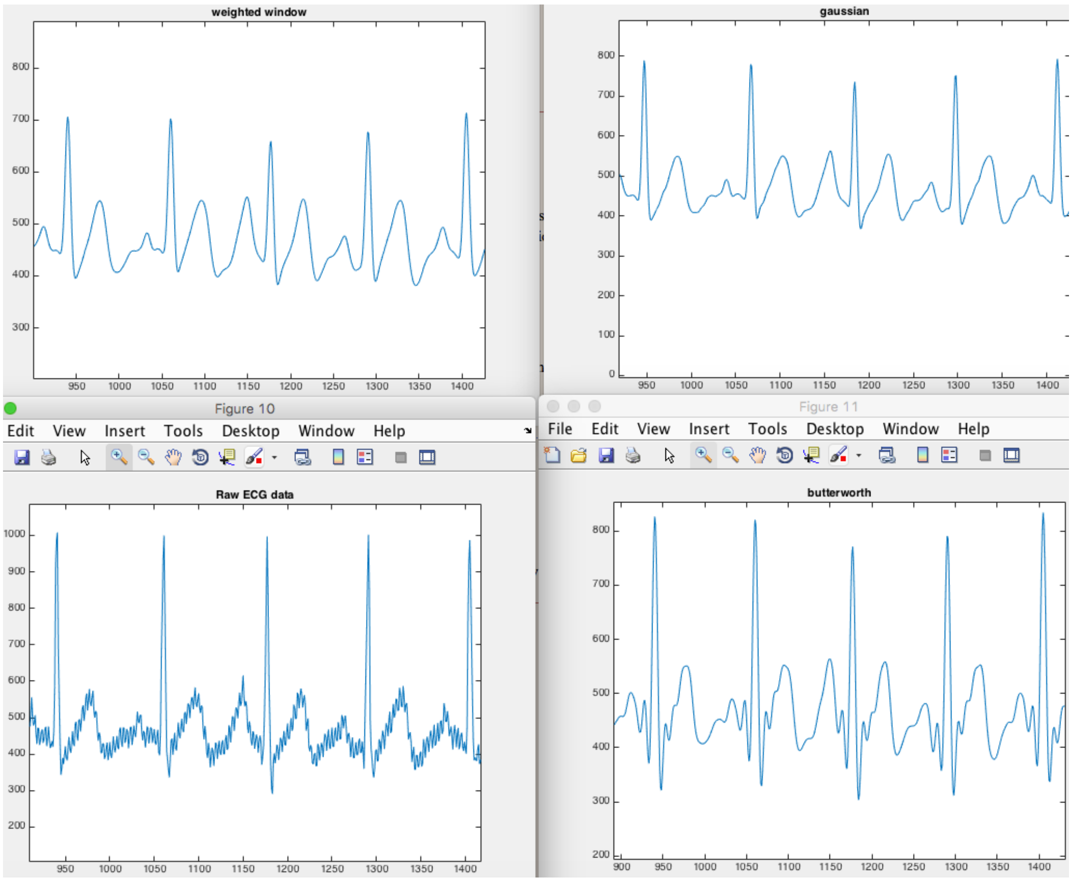

Figure 9: Filtered ECG signals and raw data

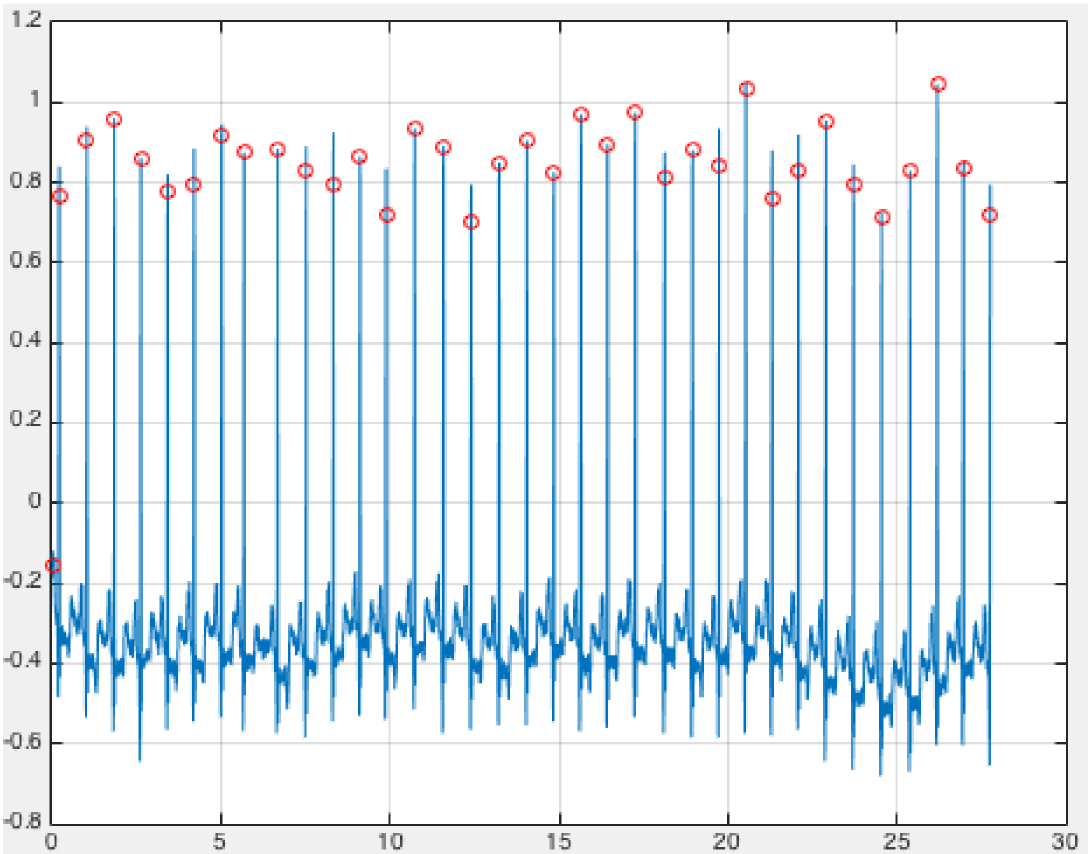

Figure 10: Imported ECG signal from physionet database

4 Plan of Remaining Work

The overall system architecture is shown in figure 11.

4.1 Further Signal Processing of ECG signals

The first phase of the project, the acquisition of ECG signals, is almost complete. The signals still must be filtered further to remove baseline wandering. Once this is complete feature extraction must take place. This will involve isolating each P-QRS-T segment, removing a distorted segments using the Euclidean distance from the mean and correcting the fragments based on heart rate. Heart rate correction will be done using a linear normalization of the extracted fiducials.

4.2 Ethical Approval

Live subjects will be used for the development and demonstration of the ECG identification algorithm. Hence, ethical approval must be sought before experimentation can begin on others.

4.3 Develop ECG ID Algorithm

The identification algorithm will use a similar model to the one discussed in Aretga-Falconi et al. [?]. Once the main features are extracted a scoring system will be developed from the time-based features. A decision will then be made on the acceptability of the score and the user either authenticated or denied. The enrollment template will be taken from a larger set of data. The aim will be to make the authentication template as short as possible to increase usability.

4.4 Testing

Tests will be performed after each stage of development. The extraction of PQRST segments can be tested by either providing or using signals with varied heart rates. Separate test cases will be developed for data acquired from the databank and data acquired from hardware. The signals from the database contain between 2 and 20 records per subject. One recording will be used for the authentication template and all other signals will be used for identification testing. Signals acquired from hardware will use a similar approach.

4.5 Make Improvements

If there is enough time, the ECG ID algorithm can be improved by extracting more features such as the radius of curvature of the fiducials as discussed by Israel et al. [?]. The hardware implementation can be improved by using a raspberry pi as the host, using python scripts to process the signals and implement the ID algorithm, instead of a PC running Matlab.

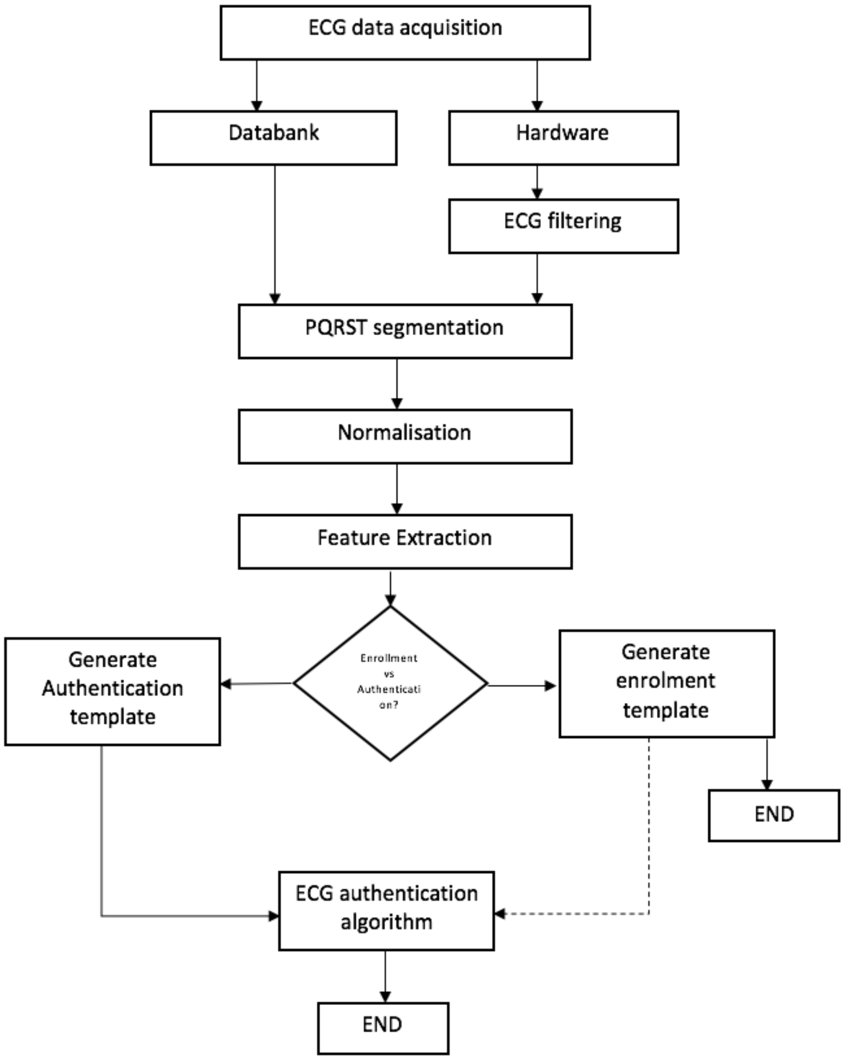

Figure 11: Overall system architecture