Ain Shams University
Faculty of Computer & Information Sciences
Scientific Computing Department

ECG Based Authentication Interface

Human Computer Interfaces (HCI)
SC_17

No	Name	ID
1	أحمد عبدالجواد رمضان عبدالجواد	20191700047
2	خالد شريف عبداللطيف عبد المجيد عزام	20191700225
3	سعد وليد سعد على ابوحسن	20191700284
4	يوسف محمد جمعه محمد	20191700785
5	عمر أحمد الشناوى عبدالحميد	20191700397

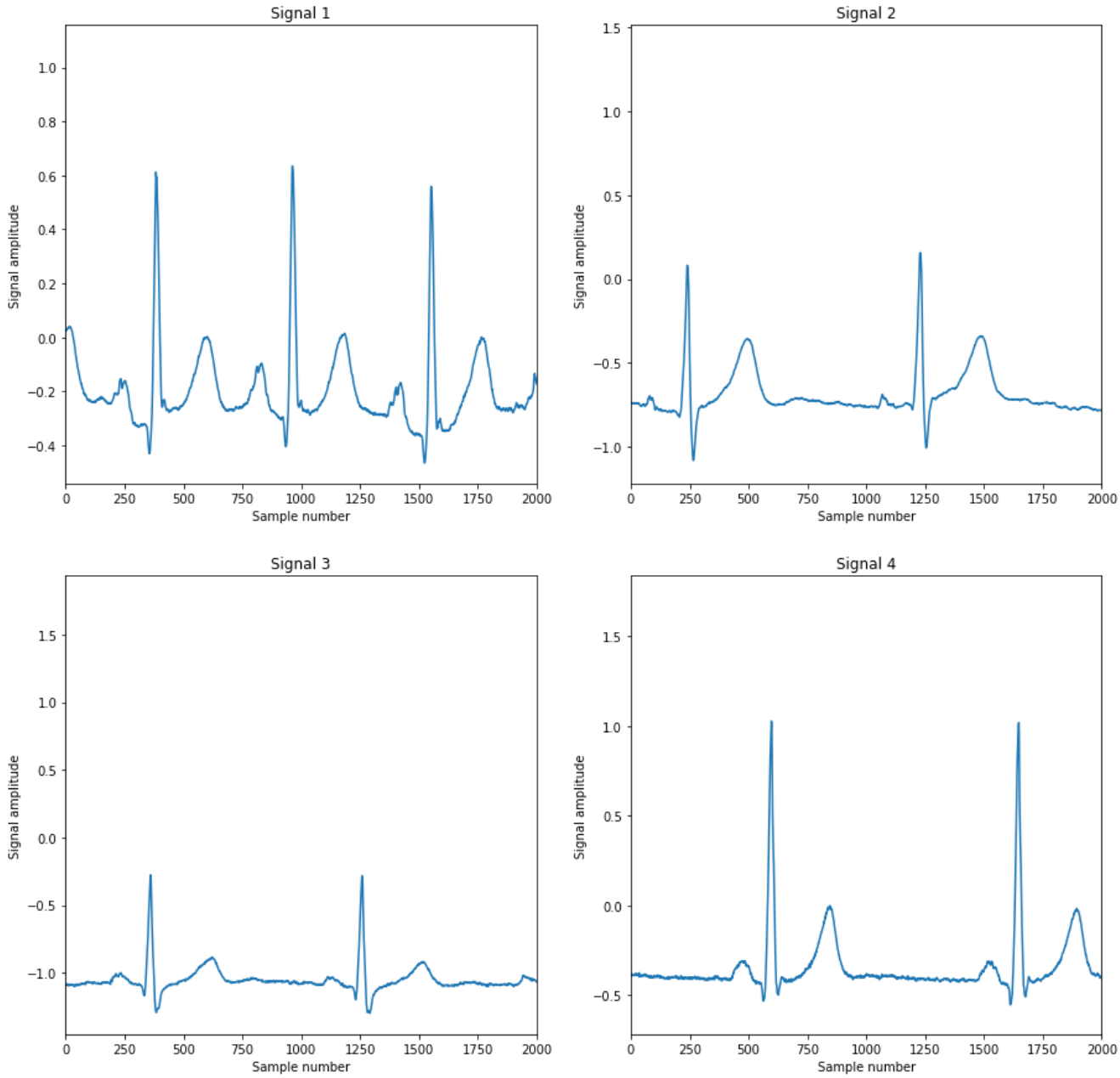
Data Preparation

- ## dataset

We have obtained 4 ECG signals for 4 different healthy subjects from PTB database. The data is split into 80% for training and 20% for testing, The description of the data is:

Number of signals	**4**
Sampling frequency	**1000**
Signal Length	**120012**
Number of channels	**15**

Signals Visualization



- # Preprocessing

We have worked on only one channel (ii).

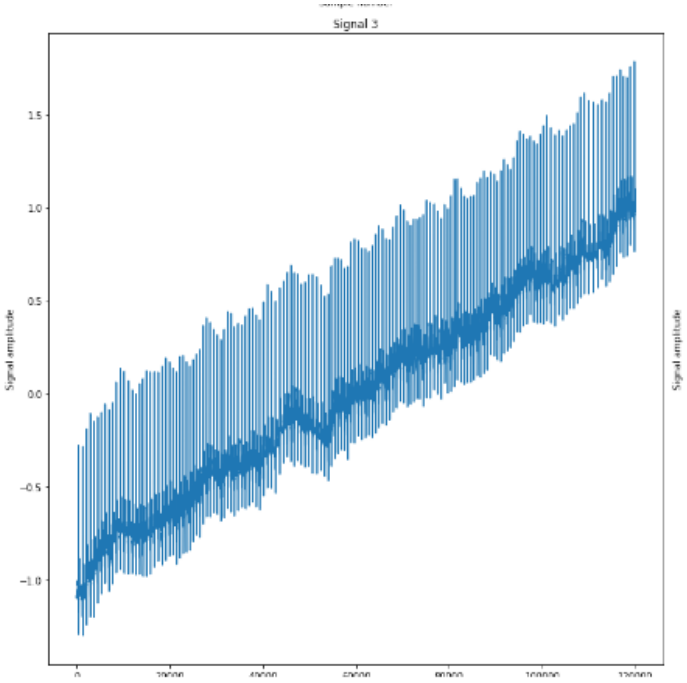
- ## Filtering

  ECG data contains noise from multiple sources (e.g., Baseline wander, Power line interface).

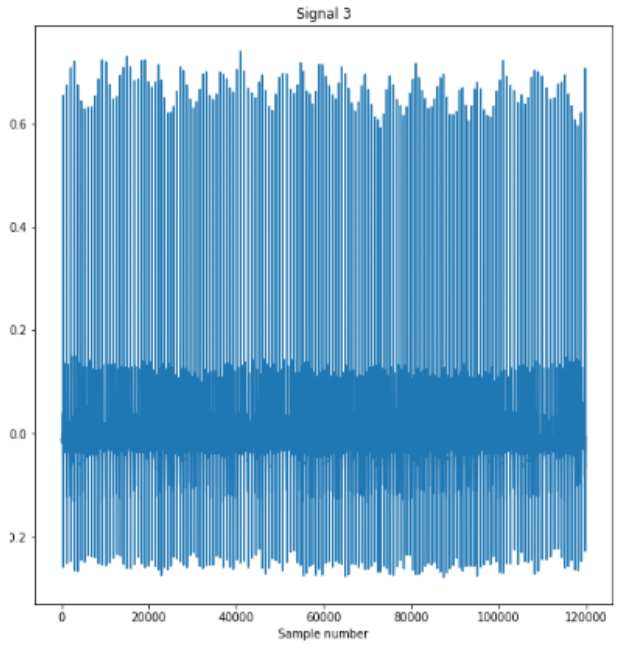
  We have used a simple Butterworth filter with cut-off frequencies between [1 – 40] Hz, which is the ECG spectrum, to remove the noise.

- ## Segmentation

  For the Non-Fiducial features, we divided the signals into segments each with length of 2000 sample and overlapping with 200 sample.

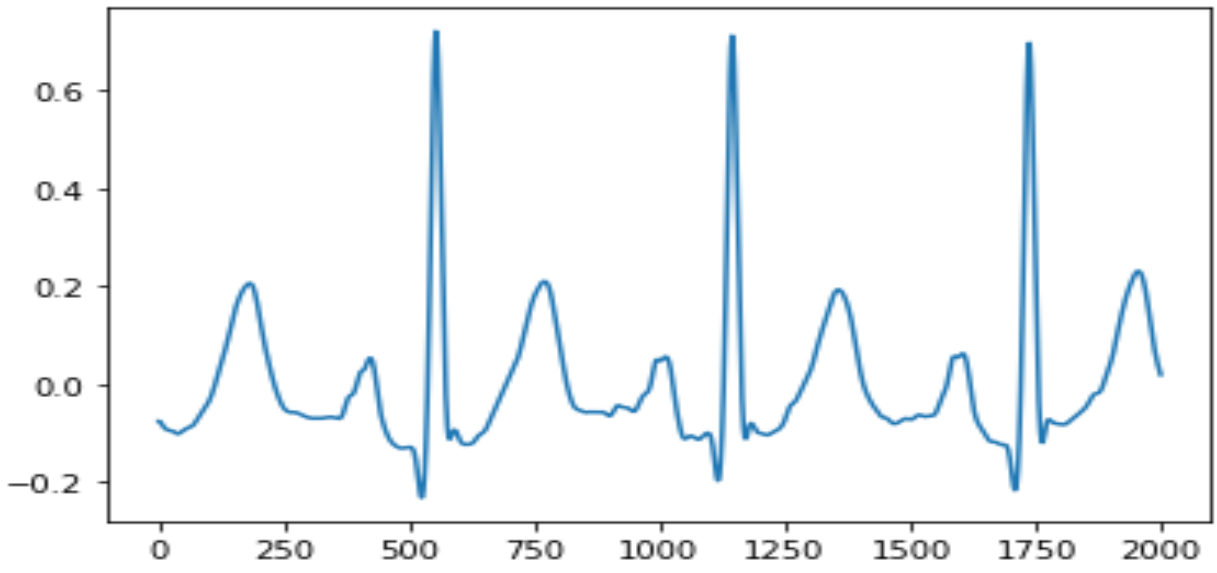


Before filtering



After filtering

This what a segment looks like:



Feature Extraction

For ECG signals there are two kinds of features to extract: Fiducial and Non-Fiducial Features.

- **Fiducial Features**
  Which represents the amplitude of the signal fiducial points: P, Q, R, S and T and the duration between them.
  11 Points to be detected the peak of each of the three Complexes QRS, P, T and the onset and offset of each of them.
  At first, we detect the R-peak by using Pan and Tompkins Algorithm which is:

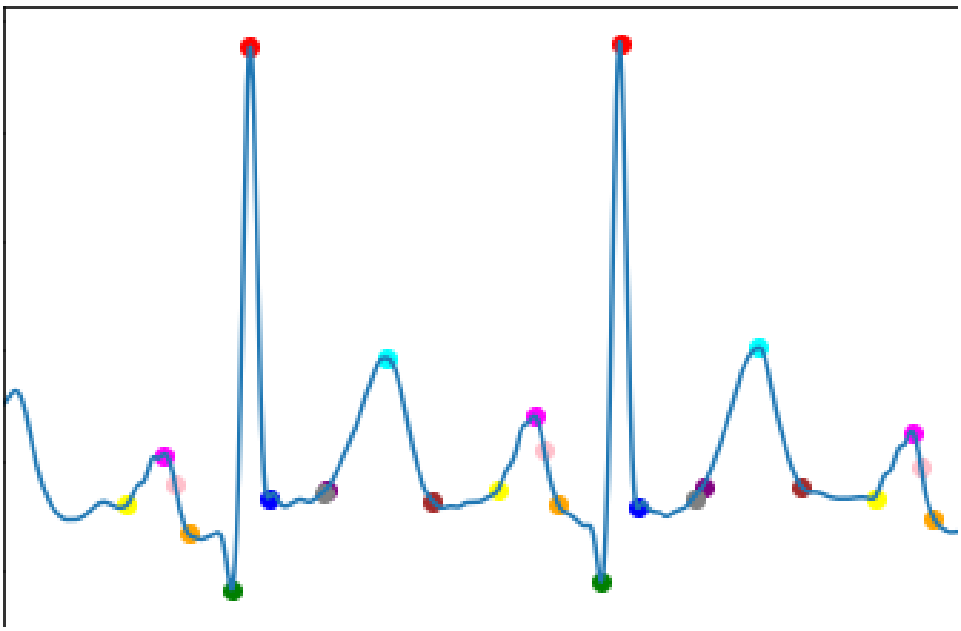
1. Band pass filter	2. Differentiation	3. Squaring	4. Moving window	5. Thresholding

Or simply find the R-peaks by detecting the local maxima on the neighborhood and setting a suitable distance between neighborhood (550 sample).

Then we can detect the P-peak by searching for the local maxima before the R-peak in a window of width 200ms ~ 200 sample (fs = 1000), by applying same method we can detect the T-peak by searching for the local maxima after the R-peak in a window of width 400ms ~ 400 sample (fs = 1000). But if the T is inverted, we will need to look for local minima not the maxima. So we can use a more general method the wing function which is defined as: $w_1 = X_{i-16} - X_i$, $w_i = X_i - X_{i+16}$ $W = w_1 * w_2$ then the T-peak will be the minimum point of the W function.

After finding the three complexes P, QRS and T., then we will find the onset and offset of each wave by using the method of minimum radius of curvature. Which is found by maximizing the value δ which is defined as $\frac{|\vec{a}\, x\, \vec{c}|}{|\vec{a}|}$ where a is the vector between the peak and the start or end of the search window, c is the vector between the peak and each point in the search window.

Here are the 11 detected fiducial points:



Using these 11 points, we can extract 36 features, which represents the amplitude and the duration between them.

We have only used 12 features out of the 36.

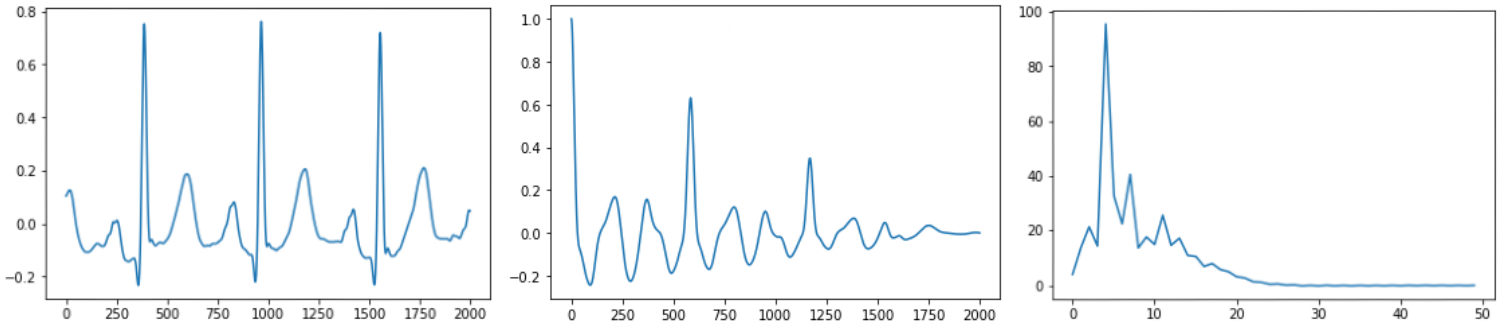
- **Non-Fiducial Features**

  The process of detecting the fiducial points is a challenging task, because of the noise and that there are no standards to exactly know where the point exists.
  So, the utilization of non-Fiducial features is needed as it represents a general pattern in the ECG signal, and it only requires the detection of R-peak or without fiducial features at all.

  To extract the non-Fiducial features, we segmented the signal into 2000 sample length long segment. And fed these segments to these methods:

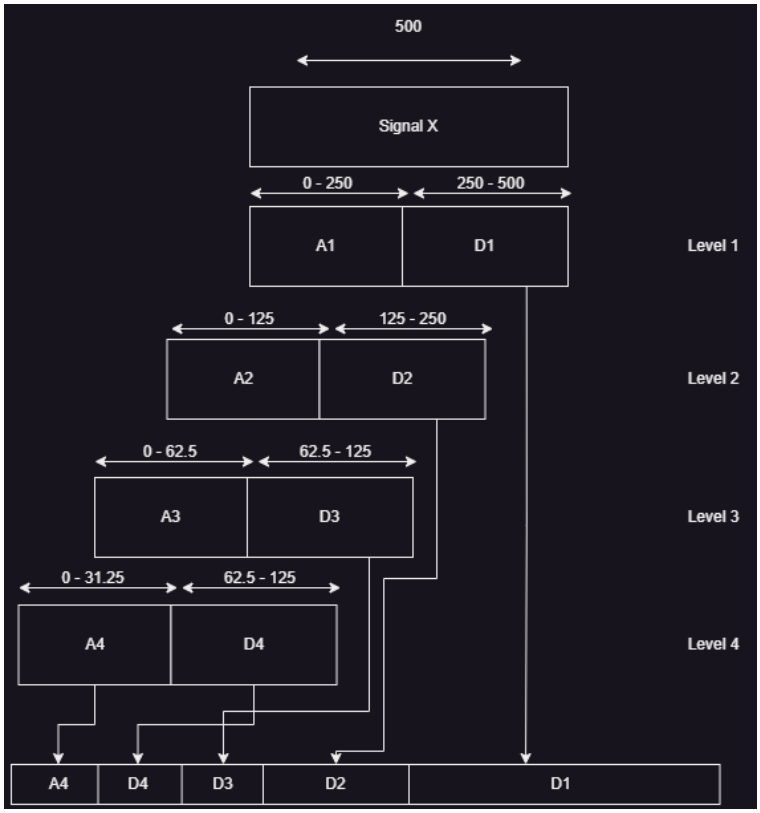
  - **AC/DCT**

    We apply Autocorrelation (AC) in each segment, and then we apply DCT on the first 400 autocorrelation coefficients to reduce the dimension, and then use that as features.



  - **Wavelets**

    We Decompose the ECG segments using a mother wavelet ('db4') with level = 4, as the fs = 1000 so on the level 4 we will get the ECG band (1 – 40) Hz and preserve these coefficients and use them as features.



Classification

We have tried multiple classifiers with each feature extraction method:

_____	Logistic Regression (liblinear)	SVM
Fiducial Features	**0.99**	**0.81**
ACT/DCT	**1.0**	**1.0**
Wavelets	**0.28**	**0.99**

We tried the Logistic Regression with different solvers like lbfg and sag with each feature extraction method, but the model didn't converge at All.

Here are the Accuracy summaries.