Report on Erasmus Placement
at the

Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris

Year 2000-2001

by

John Stephenson

MSci Physics
Imperial College, London

Detection of Isoelectric Baseline and High Frequency Noise within an Electrocardiograph Signal

Project Supervisor :
Brigitte Quenet, Assistant Professor

Laboratoire d'Electronique, ESPCI
10, rue Vauquelin
75231 Paris Cedex 05
Tel : 01 40 79 44 52

Abstracts

This study proposes several algorithms in the domain of automated electrocardiograph (ECG) analysis, regarding the detection of an accurate isoelectric baseline, and the estimation of high frequency noise. The principle method has been the determination of isoelectric regions via signal zonation and variable thresholding. Quantitative results from simulated ECG signals demonstrate high levels of accuracy for both algorithms, and are corroborated by results from the MIT signal database.

Cette étude propose deux algorithmes d'analyse automatique d'un signal ECG pour la détection de la ligne isoélectrique du cœur et l'estimation du bruit haut fréquence. L'idée essentielle est le repérage des zones d'inactivité afin de déterminer au moyen de seuils adaptatifs le potentiel de repos. Des tests effectués sur des ECG artificiels et sur la base de référence du MIT mettent en évidence l'efficacité des algorithmes proposés.

Acknowledgements

I would like to express my deepest gratitude to Brigitte Quenet whose enthusiasm, openness and approachability made this year so enjoyable and interesting.

Equally, I would like to thank Remi Dubois for his advice, knowledge, ideas and general companionship over the course of this Erasmus placement.

Additionally I would like to thank the other members of the laboratory, Pierre Roussel, Yacine Oussar and Gerard Dreyfus, who made the laboratory such an open and convivial place to work.

Finally, I would like to thank Yves Faisandier of ELA medical, whose test signals provided the cause of so many rewrites and changes of the algorithms proposed here. This report does not come close to expressing the number of ideas that needed to be rethought as a result.

All of the algorithms to be presented were programmed, researched and tested by the author throughout this year, with ideas for direction and domains of study being given by the people mentioned above. I would like to specifically thank Brigitte Quenet, Remi Dubois and Yacine Oussar for their technical help with the linear algebra, filtering methods and wavelet analysis respectively, that formed the backbone of this report.

Contents

Chapter 1. Overview of Project

In the latest worldwide study by the World Health Organisation, death from cardio-vascular disease accounted for a third of all mortalities worldwide[1]. Some of these could be avoided as they are usually preceded by characteristic signs in an electrocardiograph signal (ECG). However, the occurrence of these signs may be too exceptional to appear during a standard 10-20 second ECG record. Such signs may appear in a 24-hour signal, but it takes too much time for a cardiologist to examine each of the 100,000 beats. A system that can bring these erroneous signals to the attention of the cardiologist would allow a more accurate diagnosis to be made. Since September 2000, a research group from the ESPCI, comprising Brigitte Quenet, Remi Dubois, Gérard Dreyfus and the author, has worked in this domain of ECG analysis.

1.1 Aims of Research

This group wishes to develop a system to automate the processes in electrocardiogram (ECG) analysis and diagnosis. Whilst it does not wish to replace the essential work of the cardiologist, it aims to provide a means of long-term heart monitoring (the 24hr 'Holter[2]' record), not presently viable given the vast amounts of information present in each of signal.

There are numerous groups around the world working on different methods of automated ECG analysis, though successes have been limited owing to the immense complexity, variability and noise content of such a signal. This group aims to apply modern mathematical tools such as neural networks, wavelet analysis and hidden Markov methods (HMM), the use of whom in relation to the ECG signal is still in infancy.

This may in the future give rise to 'on-line' heart monitoring, allowing a greater availability to the system and computing power[3].

1.2 The ECG signal

The aims of the project are necessarily linked to the characteristics of the ECG waveform. These will be defined in more detail later in the report, but a summary of the terminology and characteristics of a typical ECG are included here for clarity. The ECG signal is composed of five principal waveforms, P, Q, R, S and T (see **Figure 1-1**). In addition there is sometimes a U wave between T and P waves, but its origins are uncertain. All of these waveforms are

[1] By 2010, the WHO predicts that CVD will account for the majority of all worldwide deaths. [WHO (1999)]

[2] The Holter record is made by an apparatus that the patient will wear throughout their daily activities, giving a 24 hour record of the heart's activity.[Adamec and Adamec (2000)]

[3] On-line monitors are currently available in the USA, allowing the patient to transmit an ECG record to the cardiologist by phone.

measured from the isoelectric line or baseline, which is represented by the dashed line in the figure. In addition the ECG signal contains both high and low frequency noise, with the low frequency noise termed ‘baseline wandering’.

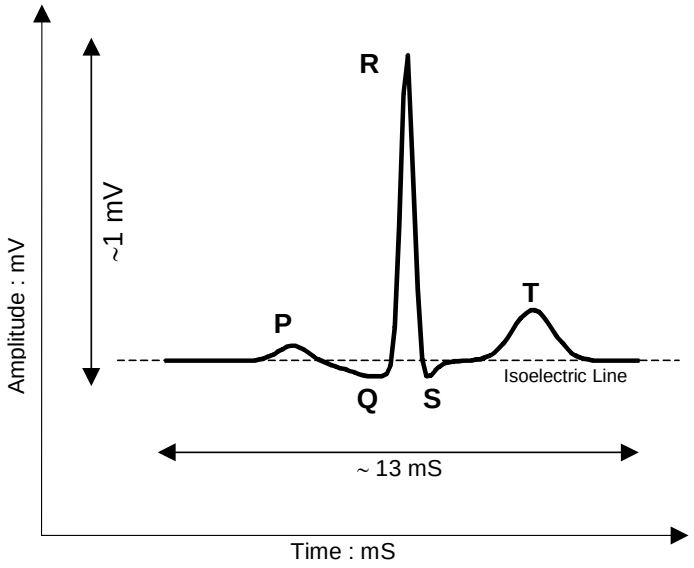

Figure 1-1 *Diagram of characteristic ECG waveforms and isoelectric line*

1.3 Group Objectives

The group’s objectives can be split into five distinct sections:

1. **Robust Detection of R waves**[4]
2. **Pre filtering and detection of baseline**
3. **Modelisation of waveforms via wavelet analysis and neural networks**
4. **Classification of waveforms via HMM**[5]
5. **Interpretation of waveform sentences and irregularities to give pathology**

1.4 Project Objectives

The principle objectives of this project and the topics covered in this report fall within the group’s second objectives. The author’s work has mainly involved the creation of algorithms to address the following aims:

[4] This has already been achieved by the group to a very high degree of accuracy via standard digital filtering and thresholding, essentially picking out the high frequency components of the QRS complex. An error of 0.6% was calculated when tested against the annotated MIT databases.[MIT-BIH (2001)]

[5] HMM classifies each waveform into P, Q, R, S or T using the probability that the wave can exist in the given position spatially (e.g. Q wave after an R wave is not possible), as well as its parameters of shape and size. [Coast, et al. (1990) ,Dugad and Desaj (1996)]

1. **The reduction of baseline wandering in an ECG signal**
2. **Detection of isoelectric line**
3. **Detection of high frequency noise interference**

To achieve these aims the following objectives were outlined by the author:

1. **Separate signal into zones**
2. **Create algorithm to identify inactive zones**
3. **Calculate a baseline to account for both baseline wandering, and actual isoelectric line, using inactive zones**
4. **Estimate amount of high frequency noise in an ECG signal, using inactive zones.**

Though not included within the project objectives, the author also wrote an algorithm that is currently used in the group's third objective, modelling the signal. This will be discussed in an introduction to wavelet analysis. Additional work was done on a spatially selective pre-filtering technique using orthogonal wavelets, which is found in appendix A1.

After introductions that will include a discussion on the mechanics of the ECG signal, as well as an outline of wavelet analysis, the report will be split into three separate topics, corresponding to the objectives outlined above. These topics are designed to be self-contained, with results pertaining to the specific topics will be included in each.

This was by nature a flexible, evolving project, with research priorities changing with the needs of the overall group project. The project objectives above give an idea of how the project was finally organised, with the majority of the work towards the end focussed on these four objectives. However, the progression through time is almost a complete reversal. Within each topic, an idea will be given as to the true progression of thought that lead to each algorithm's development.

1.5 Group Collaboration

The group is currently working jointly with Electronique Apliquée Medical Systems (ELA), with the joint aim of creating a new generation of automated Holter analysis. ELA is a commercial company and are among the current world leaders in Holter and pacemaker technologies. Input has been given to the group relating to current methods of Holter analysis (the commercially available SyneTec computer programme), as well as unbiased testing of the proposed algorithms during regular meetings. For more information, see [ELA (2001)].

1.6 Testing Methodology

The accuracy of the group's work is tested against several large databases of ECG signals. These include the MIT and American Heart Association ECG databases, as well as ECG

signals given by ELA medical. Within this project, the algorithms have been principally tested against the MIT database on a day to day basis.

1.6.1 The MIT Database

The MIT database consists of 25, 30 minute Holter ECG signals, each chosen to give specific difficulties to any automated analysis systems, including difficult waveforms and noise frequencies. Each signal has been annotated by cardiologists, allowing quantitative errors of pathology analysis to be calculated. Greater information on these can be found in [MIT-BIH (2001)].

N.B. It should be noted that in several figures involving the MIT signals, units are given for the amplitudes. However, the units in the MIT signals, are arbitrary, and should be treated as relative measurements rather than exact voltages.

Additionally, the author has tested his algorithms against several artificial signals, to allow quantitative errors in baseline detection an error analysis to be made. These include:

1.6.2 ECG Simulator

The ECG simulator was designed by Remi Dubois of ESPCI to give signals with 12 different pathology types. In addition, it was also possible to vary the heartbeat frequency. These signals represent perfect signals, allowing the author to define the levels of high and low frequency noise, and thus provide quantitative errors. Examples of each of the signal types can be found in appendix A3.2.

1.6.3 ELA Test Signal

The ELA test signal is a simulated signal, created by the representative of ELA medical systems, to specifically challenge the author's aims as outlined previously. It consists of a particularly difficult pathology (that of an infant under 1 month), as well as difficult levels of high and low frequency. Examples can be found in appendix A3.3.

Chapter 2. Introduction to ECG Signal

This chapter aims to give a general overview of the domain of this project. This is intended to give the reader a greater appreciation of the incredible variety and complexity of the ECG signal, as well as the difficulties that are faced by any automated analysis system.

2.1 Origins of the ECG Signal

An electrocardiogram (ECG) is a graphical representation of the hearts electrical activity. It gives a quantitative measure of the passage of electrical signals through the heart muscle (myocardium) causing it to contract and hence pump blood around the body. The characteristics of these signals may reveal diseases of the myocardium i.e. infarcts[6], the presence of abnormal beating cells, or difficulties in the electrical transmission between the different parts of the heart (i.e. between atria and ventricle). These kinds of disease cause incorrect contractions of the heart muscles, hence reducing the efficiency of the pump.

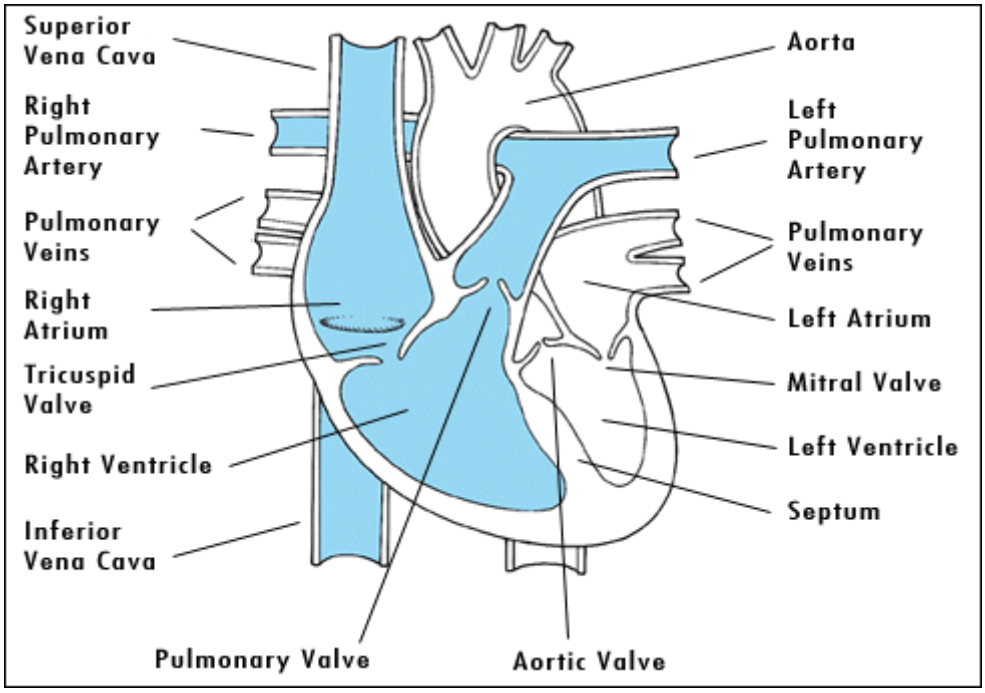

Figure 2-1 *Diagram of heart structure*

In order to understand the origins of ECG signals a certain amount of background information is required on the generation of the electrical signals and the characteristics that allow a pathology to be determined by the cardiologist.

[6] An infarct of the myocardium is a blockage of an artery in the heart, stopping blood flow within that zone, thereby causing surrounding cells to die.

2.1.1 Heart Dynamics

The heart comprises two double cavities on the left and right sides of the body, each split into a large ventricle and a smaller atrium (see figure 2-1). At each heartbeat, electrical signals flow around the heart inducing a sequence of contractions in the myocardium, causing the blood to be pushed from cavity to cavity and around the body via various major arteries. The sequence commences via the sino-atrial (SA) node at the top of the right atrium, emitting electrical impulses at regular intervals (usually between 60 and 100 beats per minute). This causes the right atrium to contract pushing blood into the right ventricle via a one-way valve. The right ventricle then contracts pushing blood down the pulmonary artery to oxygenating processes in the lungs, before returning to the left atrium via the pulmonary vein.

The left side simultaneously undergoes the same processes as the right, thereby pushing the now oxygenated blood (from the previous cycle) from the atrium into the largest chamber, the left ventricle, from where blood is pushed to the rest of the body. Nomenclature dictates that the period of contraction is termed the systole, whilst that of dilation, where blood fills the atria after contraction, is termed the diastole. [DeBakey and Gotto (1997)]

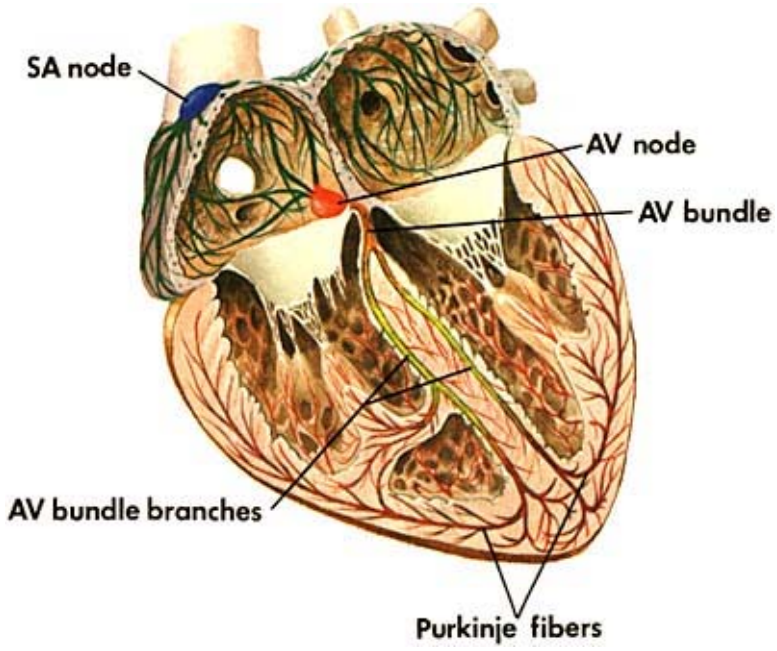

Figure 2-2 *Diagram of heart interior, showing electrical pathways*

2.1.2 Cause of conduction

The sequence of electrical propagation is initiated by impulses from the SA node (see figure 2-2), though it has not been made clear how the wave is able to propagate through the heart. This is essential in understanding the characteristics of the ECG signal.

The cells of the myocardium can exist in two electrical states, depolarised (during contraction phases) or polarised (during rest). During the latter state, the interior of the cell is negatively charged, whilst the exterior (cell surface) is positive. When the cell is stimulated (electrical impulse from the SA node, or surrounding cells), the interior of the cell becomes positively

charged and the cell contracts. Each cell then influences its neighbouring cells, causing a wave of current to flow (the progression of a wave of positive charges), commencing from the SA node, and spreading throughout the heart. This process is termed depolarisation and drives the contraction sequence of the myocardium .In addition to this cell to cell transfer of charge, there are several distinct paths through which the wave propagates extremely quickly (similar to an electrical cable), thereby ensuring the simultaneity of contractions (see figure 2-2).

It is this electrical wave that is interpreted in the ECG signal. Each of the waveforms in figure 2-3, correspond to the depolarisation or repolarisation of different areas of the heart. As such, their shape and characteristics allow a non-invasive look at how the heart is functioning. [Dubin (1999)]

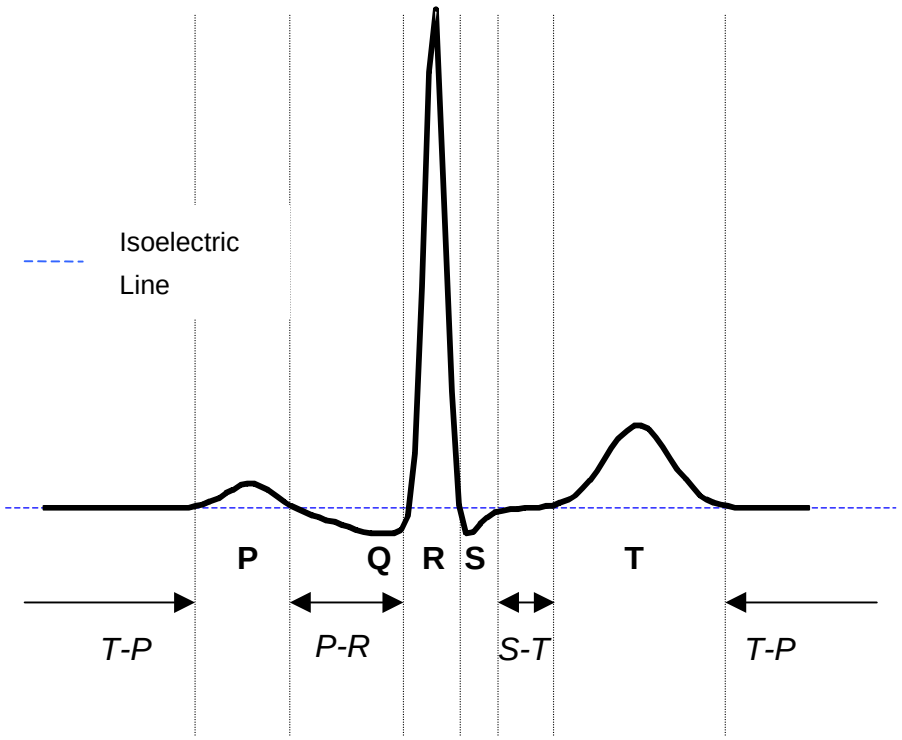

Figure 2-3 *Characteristic waveforms and segments of a normal ECG signal*

2.1.3 The Electrocardiograph

These waves of impulses can be picked up from the surface of the human body and measured using essentially a very accurate galvanometer, the electrocardiograph. The torso and members of the body are connected to the electrocardiograph via a series of positive electrodes called leads. Each lead also has a negative reference electrode, which it may share with another lead, depending on the lead type.

By looking at the potential difference between the positive and negative electrodes in a lead, an idea of the direction of the electrical wave is given. Hence the electrodes are placed around the body in such a manner that a map of the entire heart is given. The electrodes are orientated such that when a wave is coming towards the positive electrode, a positive (upward) trace is

made on the ECG paper, and vice versa. Each of the leads records essentially the same signal of PQRSTU, though certain of the waveforms are accentuated by the individual orientation of each lead. The number of leads varies with the system with the modern hospital electrocardiograph consisting of 12 leads, thus mapping 12 different axis or 'derivations' of the heart, whilst the 'Holter' system has usually only 3. See appendix 3.1 for an example of the different ECG signals from each of the 12 derivations in a standard hospital electrocardiograph.

2.1.4 The Holter Record

The group is directed specifically toward Holter or ambulatory electrocardiography. Here an ECG signal is taken throughout a 24-hour period, via a small device that the patient will wear throughout their daily activities. The main aim of Holter recording is to capture the occurrences of abnormal electrical behaviour in the heart, that may occur very infrequently and therefore not captured by the standard ECG recorder. Norman Holter initially designed this system in the 1960's. [Adamec and Adamec (2000)]

2.1.5 PQRSTU Origins

Using this information of cell electrical activity, and the way it is measured, an understanding of the important segments and waveforms of the ECG can be established (see figure 2-3).

- **Isoelectric Line**

The isoelectric line or baseline is the natural zero of the signal from which the cardiologist will measure the amplitudes and the start and the end of all of the waveforms and segments to follow. It is measured from the region of inactivity between the T and P wave, and is sometimes seen in the PR and ST segments.

- **P Wave**

The P wave represents the depolarisation and subsequent contraction of the two atriums, following an impulse trigger from the SA node.

- **PR segment**

Following the end of the P wave, there is a period of pause in the heart. Electrically this corresponds to the time the wave spends in the AV node, before passing to the ventricles. Physically this pause allows blood to refill the ventricles. The signal is then directed towards the base of the ventricles, propagating rapidly via His's bundle (sometimes termed the AV bundle), and its left and right branches.

- **QRS complex**

Essentially the QRS complex is representative of the depolarisation of the ventricles. The R wave is defined as the first positive deflection in the QRS complex, with the Q wave defined as the first negative deflection before the R wave, and the S wave as the first negative deflection after the R wave. The lead derivation being studied, and any myocardial infarcts present determine the presence of a Q wave. The R wave and the S wave then represent the passage of the depolarisation from His' bundle to the myocardial cells and subsequent

ventricular contractions. This process is via many very fine fibres (Purkinje fibres) and as such is transmitted very rapidly, as is seen in the high frequency of the QRS complex. (see figure 2-2).

The form of the complex varies greatly depending on the lead chosen and its orientation to the heart. For example, in some the S wave may be larger than the R. In this report, we shall stick to the orientations of lead I as shown in figure 2-3.

- **ST segment**

After the QRS, there is another pause in the ECG where there is no electrical wave in the myocardium. This period is extremely important in the diagnosis of numerous pathologies with [Dubin, 1999 #13] listing 5 different heart diseases that are evident from its study.

- **T wave**

This last waveform represents the repolarisation of the ventricles, allowing the cells to regain negative charge, and therefore equilibrium, ready for the next heartbeat. It does not correspond however to any physical motion in the heart. N.B. a wave for the repolarisation of the atriums is present, but it is insignificant in relation to the QRS and hence invisible on the ECG record.

- **U Wave**

The U wave is sometimes present after the T wave. Its origins are unknown, though it is believed to result from the repolarisation of the Purkinje fibres

In summary, information of the physical motion (contractions) of the heart is gained only from certain parts of the signal, with the remainder giving indications of how electricity is flowing in the myocardial cells. [Dubin (1999)]

2.2 Introduction to Cardiology

There is an innumerable amount of information present in the ECG signal, all of which can be used to diagnose a pathology.

There are 3 particular areas of interest that we shall discuss here: Rhythm, Hypertrophy and Infarcts. All of these areas have similar importance. (NB. There is a 4th, Axis, which we can ignore here for the moment as it deals with multi-derivations, and what can be seen by comparing them.

2.2.1 Rhythm

The SA node, emitting at a frequency of 60-100 impulses per minute, regulates the frequency of the heartbeat. Deviations from these frequencies are termed bradycardia if below 60 beats per minute (bpm), and tachycardia if above 100bpm. In addition to the SA node, there are other pacemaker potentials which are able to stimulate the heart into the contraction sequence,

and are defined by the region of their origin i.e. atrial, junctional and ventricular. If the SA node is not functioning properly, one of these pacemakers will act in place, giving a characteristic heartbeat frequency and ECG shape. They are all capable of beating very rapidly at around 150-250bpm, in urgent situations or in some specific pathologies.

All areas of the heart have a certain pacemaker potential, though when they become active, they are termed an ectopic centre. Erroneous ectopic centres, also cause deviations from the normal SA node rhythm, termed arrhythmia, where pacemaker potentials cause incorrect contractions, reducing the efficiency of the heart.

Arrhythmia can be sub divided into 4 sections, and are defined again by the position of the ectopic centre (atrial, ventricular or junctional):

Irregular Rhythms:
These are generally caused by the irregularity of an ectopic centre controlling the hearts rhythm, and include variations in the SA node impulse frequency (demonstrated by an irregular rhythm with identical P waves), a wandering pacemaker where the origin of the initial impulse varies over time (characterised by varying P waves), and most dangerously, atrial fibrillation (see figure 2-4). In the latter case, the heart is stimulated by rapid signals from several ectopic centres in the atria, causing the atria to ‘quiver’, producing blood clotting and hence possibly leading to strokes.

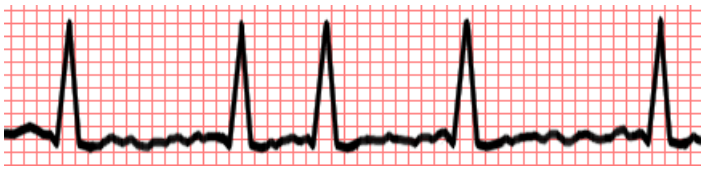

Figure 2-4 *Atrial Fibrillation*

Escape Beats and Extrasystoles:
These are additional contractions provoked by the discharge of an ectopic centre. In an escape beat, a pause in impulses from the SA node provokes an ectopic centre into discharging, and fail-safes against SA node malfunction. An extrasystole is a premature beat given by the sudden activity of an ectopic centre. In the case of a premature ventricular contraction (PVC) the electrical wave does not follow the normal fast route via the His bundle and branches to the ventricles, and gives a much larger QRS complex (see figure 2-5). PVC’s can be found isolated or else in series, and can be further complicated by various centres in the ventricles discharging, each producing a different QRS trace owing to the differing electrical paths taken. A grave situation occurs if the PVC falls on the T wave, and can invoke a very high ventricular contraction frequency.

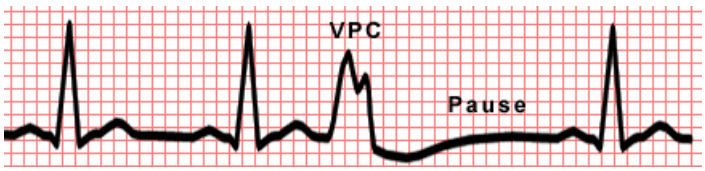

Figure 2-5 *Ventricular Premature Contraction*

Very High Frequency Contractions

The gravest pathologies occur if the ectopic centre begins to discharge at very high frequencies. They are defined in increasing severity as: tachycardia paroxysmal (150-250bpm), flutter (250-350) and fibrillation (350-450), and are further sub-divided by the position of the ectopic centre causing the rapid stimulus (atrial, junctional and ventricular). These types of pathology are incredibly dangerous for the patient, with the pump hardly functioning, and are generally precursors to sudden cardiac death. They are immediately evident in an ECG trace (see figures 2-6 and 2-7).

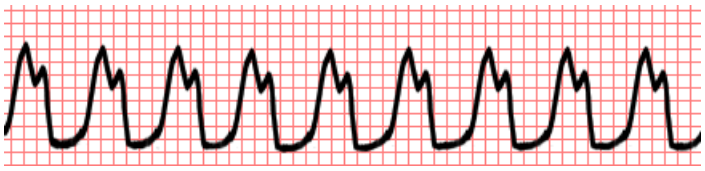

Figure 2-6 *Ventricular Tachycardia*

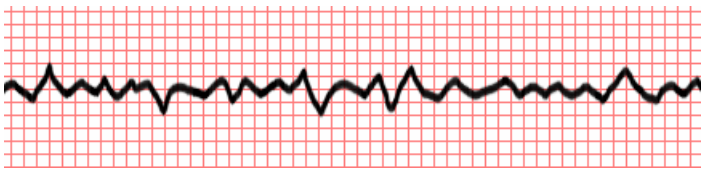

Figure 2-7 *Ventricular Fibrillation*

Cardiac Blocks

This final section concerns blockages in the passage of the electrical signals around the hearts normal conduction pathways. Blocks can occur at the SA node, the AV node, or in the His bundle and subsequent bundles. In the first, the discharge from the SA node is paused and is regained, distinguished by a long pause between QRS complexes. There are three types of AV block, distinguished by orders I II and III and increasing severity. The first order is characterised by an extended PR segment (see figure 2-8), the second by a missing QRS complex within a sequence of normal heartbeats, and the third, representing a complete block, where there are no normal QRS complexes, with the ventricular contraction given by a ventricular pacemaker. In the case of blocked branches, the two ventricles no longer contract simultaneously, and are evident in a longer QRS complex with two peaks representing the separate depolarisation of the left and right sides.

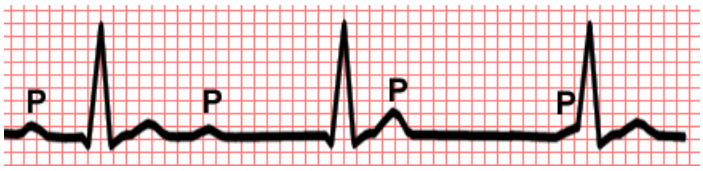

Figure 2-8 *AV block I*

2.2.2 Hypertrophy

The hypertrophy of a heart cavity (atria/ventricles) signifies an increase in the thickness of that cavities muscular wall. A hypertrophy can be seen on an ECG trace by virtue of variations in the shape and amplitude of the waveforms. For example, in the case of left ventricular hypertrophy, there may be evidence of a much larger R wave and commonly an inverted T wave, evident in a deformed ST segment. In the case of atrial hypertrophy, the P wave is deformed, perhaps having a positive and a negative part. For a full idea of cavity hypertrophy, it is necessary to study other lead derivatives.

2.2.3 Infarct

A cardiac infarct describes a blockage of an artery surrounding the heart, preventing blood flow to that region. There are various characteristic signs, including the presence of a Q wave, the elevation or lowering of the ST segment, or the inversion of a T wave, but again they are more or less visible, depending on the lead derivation.

There are in addition, other pathologies that do not fit into the above sections. These will include waveforms from artificial stimulation (i.e. an inserted electrical pacemaker), changes in electrolyte concentration in the heart and the effects of various drugs on the hearts stimulation. [DeBakey, 1997 #16; Dubin, 1999 #13]

In summary, there are several major points to come out of this introduction with reference to the remainder of the project:

- The wide variability of ECG signal
- Importance of recognising each waveforms shape and amplitude in determining the pathology
- The need for a filtering system that does not deform any of these waveforms and therefore lead to a false diagnosis
- The importance of the isoelectric line in determining the pathology type
- The small differences that can completely alter a pathology diagnosis
- Variations are evident, even in normal ECGs, as each patient has their own particular ECG signatures.

Chapter 3. Introduction to Wavelet Analysis

This brief introduction is designed to give a background into the methods that the research group are using to analyse and interpret the ECG signal. It is also necessary precursor to the understanding of why the calculation of the isoelectric line is so important.

3.1 Modelling the signal

The aim of the group is to automate a process which follow as precisely as possible the decision making process followed by the cardiologist in their analysis of a patient's. With this in mind, the group aims to decompose the signal into specific, localised waves, termed wavelets, that represent each of the areas of interest in the given ECG signal. [Vidakovic and Müller (1991)] It is then possible to match these waves with the properties of the six waveforms P, Q, R, S, T and possibly U, using advanced probability studies (HMM's), and thus identify the pathology. This initial process of modelling is termed wavelet analysis. There are several distinct steps:

1. **Choice of wavelet**
2. **Creation of wavelet library**
3. **Wavelet selection by Gram Schmidt orthogonalisation**
4. **Wavelet addition to recreate signal**

Treating these out of order, we commence with 3, wavelet selection. Here we wish to choose from a large library of wavelets, the ones that best model the signal. Gram-Schmidt orthogonalisation essentially converts this problem to one of linear algebra. In this method, the signal and each wavelet (**Y** and $\mathbf{W}_i$ respectively), are treated as column vectors of dimension n, where n represents the number of points in the heartbeat cycle being treated. . Hence $\mathbf{W}_i = [W^1_i \;\; W^2_i \;\; W^3_i \;\ldots\; W^n_i \;]^T$, where i = 1,2…. N. Each of these vectors are normalised[7], such that their total length is 1. Using the N wavelet vectors we are able to create a matrix A representing a 'library' of wavelets such that:

$$A = \left| \boldsymbol{W}_1 \;\; \boldsymbol{W}_2 \;\; \boldsymbol{W}_3 \;\; \ldots \;\; \boldsymbol{W}_N \right| \qquad \textit{Equation 3-1}$$

Hence if the library 'A' contains n wavelets all linearly independent (rank (A) = n), the wavelets will represent a non-orthogonal basis of a vector space with dimension n. As such, any vector with dimension n, can be recreated using this basis, (this is analogous to using i, j and k in three dimensional space). This is achieved by multiplying each wavelet by the projection of the vector **Y** onto that wavelet. These projection values are given by:

[7] To normalise each vector $\mathbf{W}_i$ is divided by $\mathbf{W}_i\mathbf{W}_i^T$, which is a scalar that represents its length in the vector space.

$$P_i = \boldsymbol{W}_i^T \boldsymbol{Y}$$ (i.e. the scalar product of the two vectors). *Equation 3-2*

P_i is thus representative of the angle between the two vectors. As this library is a basis of the vector space, we can write:

$$\boldsymbol{Y} = \sum_{i=1}^{N} P_i \boldsymbol{W}_i$$ *Equation 3-3*

Thus the sum of all the wavelets and the projections onto them, exactly represent the vector **Y**.

3.2 Gram Schmidt Orthogonalisation

The principle idea of this process is to order the wavelets according to how well they 'model' the signal, i.e. they represent it. These can then be added together with the correct projection coefficient as shown previously. The aim is to only add a few of these wavelets, (the most pertinent), so as to pick out only the important features of the signal Y. It is insufficient however to order the wavelets by their scalar product, as this will choose wavelets, all around the same sort of shape, hence modelling the major feature of the signal to a high degree of accuracy, leaving other features till later. This is resolved by projecting the wavelets and the signal, into successive orthogonal subspaces.

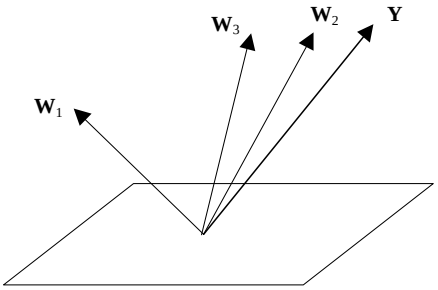

Figure 3-1: *Geometrical representation of three wavelets, W_i and the ECG signal Y, in vector space with dimension N.*

In figure 3-1, the signal **Y** is shown graphically, with three of the N wavelets. It is evident that the closest wavelet to the signal (i.e. the vector with the largest scalar product) is $\mathbf{W}_2$.

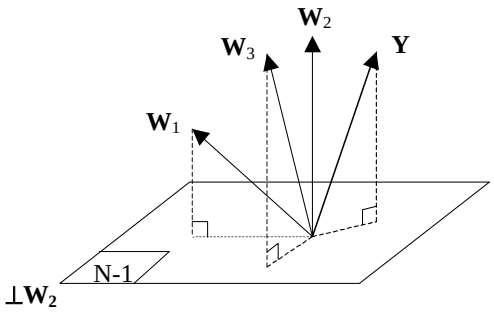

Figure 3-2: *Projection of Vectors into subspace orthogonal to the first chosen wavelet W_2.*

As such W_2 is selected as the first wavelet, being the closest vector to the signal. The next feature of the ECG signal is then found by projecting all of the vectors into the subspace that

is orthogonal to this first chosen vector (see figure 3-2). This is done using an algorithm by Gram-Schmidt, thus creating a new basis of vectors with rank of N-1. The wavelets are then re-normalised to remove any bias due to length of vectors.[8][Lopez (2001)]

Hence within this new subspace, orthogonal to vector W_2 (see figure 3-3), we are able to repeat the process, this time finding the projected wavelet that is closest to the projected vector $\mathbf{Y}^{\perp W2}$. In this way the second wavelet is chosen, by virtue of the largest scalar product $P_i^{\perp W2}$, calculated between vectors $\mathbf{Y}^{\perp W2}$ and $\mathbf{W}_i^{\perp W2}$. As can be seen, it is no longer necessarily $\mathbf{W}_3$ that is the closest vector in the new subspace, with all the features in the direction of W_2 now accounted for.

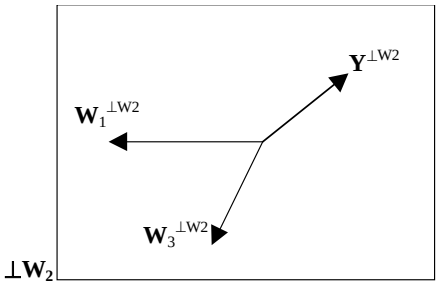

Figure 3-3 : *Representation of new vector subspace with vectors projected into subspace orthogonal to W_2.*

The vector space orthogonal to this second chosen vector is then found, thus reducing the rank of the vector space again by one and so on, iterating until either all of the wavelets have been placed in order, or a desired accuracy is achieved.

In fact, for modelling the ECG signal (assuming a correct choice of wavelet type in the beginning) only six or seven wavelets are needed to correctly give a representation of all the major waveforms. A small number of wavelets is essential as otherwise the wavelets will not be a useful representation of the desired features, giving problems in the later stage of waveform recognition by HMMs.

3.3 Choice of Wavelets

Returning now to no.1 in the list, it becomes evident that the choice of wavelet is paramount in reducing the number of wavelets selected, and thus the accuracy with which features are distinguished. After numerous studies, the group has chosen a Gaussian for the wavelet shape. The most important feature or these wavelets, with reference to baseline detection, is the fact that they most be either all positive, or all negative. As such, they model the waves starting from a zero point i.e. the isoelectric line. This is an important point, relevant in many Gram-Schmidt systems, where the choice of zero-point is essential.

[8] In fact, the group are currently working with a Gram-Schmidt method that does not normalise the vector after each orthogonalisation, which has lead to improved results. The mathematical theory around this is still being studied.

3.4 Library Creation

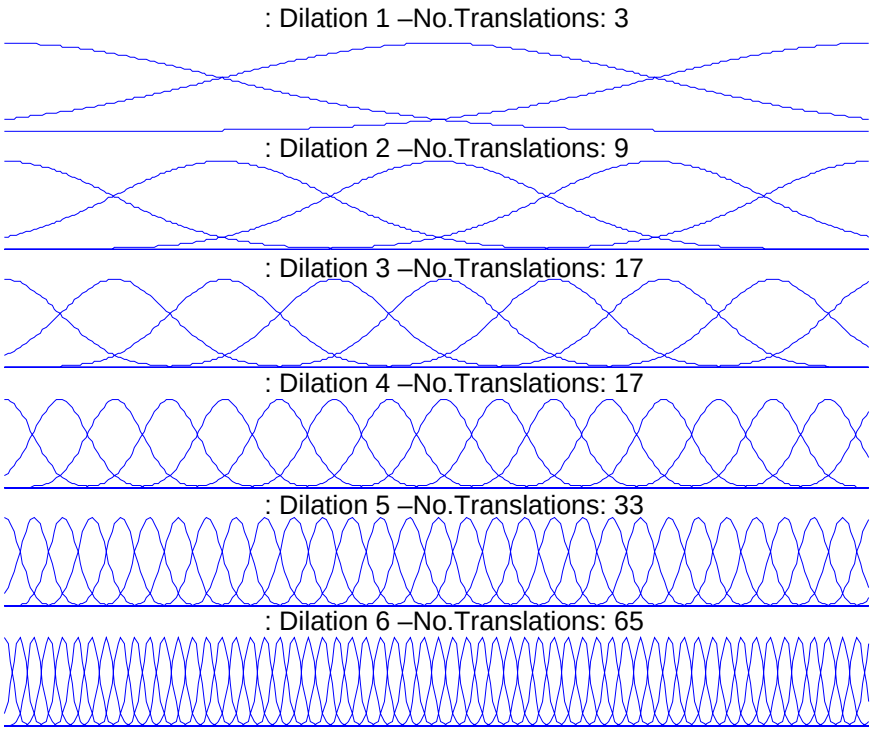

Figure 3-4: *Library of Gaussian Wavelets (each level represents a different dilation)*

Once the wavelet type has been defined, the library is created. It is created by first dilating the wavelet, and then translating it laterally. Hence, the more the wavelet is dilated the higher the number of translated wavelets for that particular amount of dilation see figure 3-4. As the wavelets are not orthogonal, it is not essential that the vectors in the library are linearly independent, or a basis of the vector space; hence, there are an unrestricted number of dilations and translations possible. In this case there are 122 wavelets created, each with a vector of length 300.

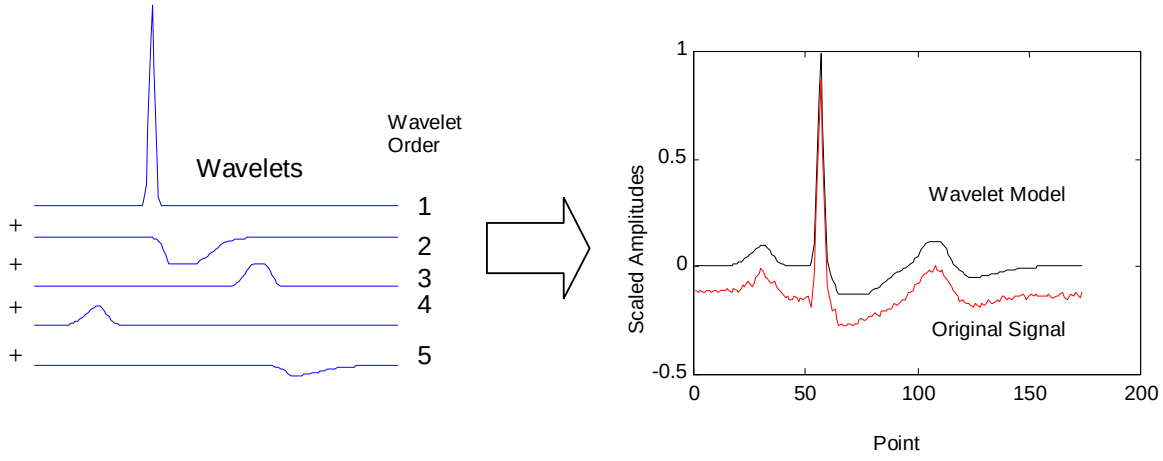

Figure 3-5: *Wavelet analysis with 5 wavelets chosen and modified, added to give a correct model*

Figure 3-5, gives an example of this group's work with wavelets. From the library of Gaussian wavelets, as described above, the 5 most pertinent wavelets have been selected. The parameters of the wavelet are then altered, so that they best fit the part of the signal they are modelling. This is achieved by a neural network learning process, that considers the wavelet as the combination of two half Gaussian curves, connected by a flat section, termed 'bumps'. The neural network will adjust these parameters until the best possible representation of the waveform is given. These are usually very small changes. (These adjusted wavelets are shown in figure 3-5).

It is these adjusted wavelets that are then passed to the HMM algorithm, which is able to identify the waveform by the following parameters: slope of left half Gaussian, length of flat section between Gaussians, slope of right half Gaussian, relative amplitude and the wavelets position temporally relative to the other waveforms. For example an R wave will have parameters such that, both left and right half Gaussians will have a very sharp slope, there will be no flat section between the two half Gaussians, the relative amplitude will be high and it will usually be preceded by a Q wave and followed by an S wave.

3.5 Author's Contribution

The author's work in this domain, was the writing of a time efficient algorithm for the process of wavelet selection by Gram-Schmidt orthogonalisation. It was realised that it was not necessary to re-orthogonalise the library after each iteration, or to calculate the entire orthogonal basis at each step. This Matlab Program can be found in appendix A4, and is the program that is currently used by the group to order the Gaussian wavelets. Help was given initially regarding the required linear algebra interpretation, though the programming and subsequent improvements were made autonomously.

Chapter 4. Zone Detection

This section deals with the method of splitting the ECG signal into zones. The author has derived all of the remaining algorithms (baseline detection and noise level estimation) from this process.

4.1 Introduction

The aim is to separate the given ECG signal into a series of zones, and give each a number representing its activity ('importance value'). From these importance values, decisions can be made over which zones represent where the heart is active (to be used in analysis), and those where the heart is inactive. This decision process will be covered in the next chapter on baseline detection.

4.1.1 Criteria

Several important criteria must be respected in this algorithm.

1. **Number of zones**
   In a normal ECG signal, with all waveforms present, approximately 10 zones will be required to separate all the waveforms and segments in the signal (P, Q, R, S, T, U, P-R, S-T & T-P). If there are many more than this, there is increased risk that the important waveforms are split, leading to erroneous results. In the alternate case, with too few, each zone will encapsulate too much of the trace, and leave too little of the signal for baseline or noise analysis.

2. **Zones Intelligently chosen**
   Each zone must represent a signal characteristic waveform or segment of the signal, so that we are able to treat each as an entire whole, and not as a series of zones.

3. **The calculation time is not too great**
   This constrains the method of filtering and region selection. This is connected in part to the total number of zones selected and the sampling frequency of the trace. N.B. the eventual aim is to perform the analysis in real time.

4.1.2 Advantages

The great advantage of this zonation method is the manner in which the regions can be used afterwards. As these zones are only time dependent, defined by the starting and endpoints of waveforms and segments, they can be applied to a signal no matter what has been done to its amplitude. As such, the regions can be located accurately using a filtered signal, which can then be applied to the original ECG to calculate the actual baseline and noise values.

This feature is important, as the frequencies of the waveforms can be very similar to those of the noise. Thus in any filtering method, there is a risk of deforming the ECG signal.

4.1.3 Low frequency Filtration

With the above in mind, it becomes possible to apply some standard digital filtering to the signal, to remove the underlying low frequency noise. This range of frequencies, from the lowest frequency to around 5Hz, is present in the signal due to the alteration of the heart axes that the lead is measuring. This is caused principally by respiration altering the position of the heart within the body, and other physical movement of the electrodes on the patient's skin.

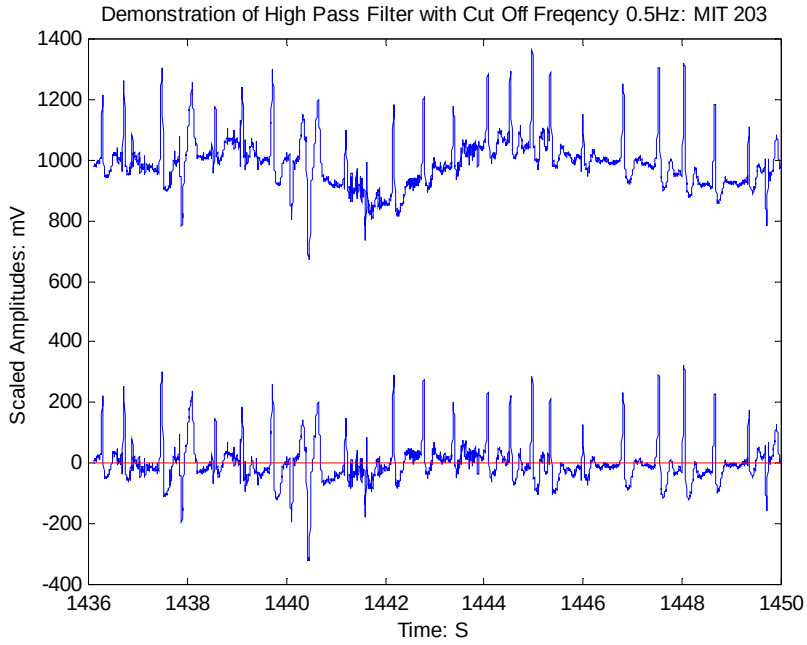

Figure 4-1: *Reduction of baseline wandering via a high pass filter, with cut off frequency of 0.5Hz*

As a first step to reduce this effect, the signal was passed through a high-pass filter, with a cut off frequency of around 0.5 Hz. This does not greatly affect the shapes of the waveforms, but does allow a much more accurate 'importance value' to be given to the zones found, as will be seen later (This is demonstrated in figure 4 - 1). The actual filter is a high pass Butterworth, and was designed using the Matlab signal processing toolbox. Its frequency response can be seen in figure 4-2.

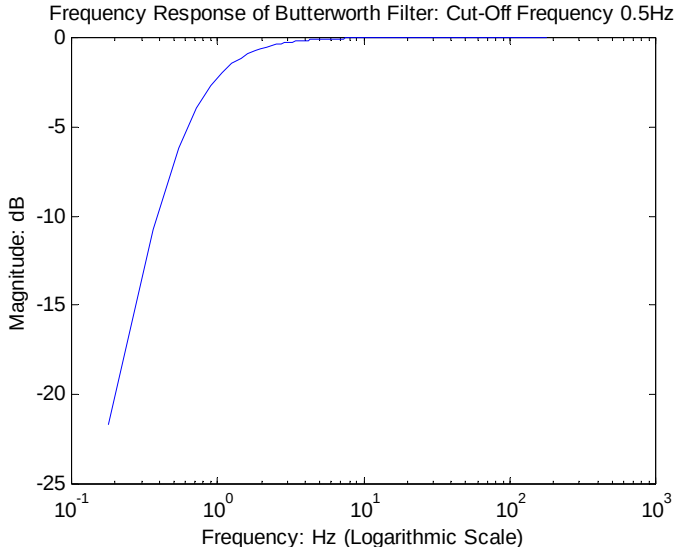

Figure 4-2: *Logarithmic representation of the frequency response from a Butterworth filter*

This figure demonstrates that frequencies below 0.5Hz are removed from the ECG signal's frequency spectrum.

4.2 Initially proposed system

Over the course of the project, the method used to segment the signal has been continuously modified and improved. Some of the important steps leading up to the final method of zero-crossings are outlined below, highlighting important ideas gained from each method along the way, and the subsequent improvements incorporated into the final system.

4.2.1 Triangle Fit of Maxima and Minima

This initial method was developed for a proposed filtering method using wavelets (see appendix 1) and also incorporates a method to distinguish between inactive and active zones. The aim was to select the region between two R waves, most likely to correspond to noise (thus allowing precedence to be placed on those wavelets modelling waveforms rather than noise during wavelet selection). It was realised that this region also allowed the best estimate of the isoelectric line, which lead to further work on determining the baseline of a signal.

The first step, as with subsequent methods, was to split the ECG into manageable sections, achieved by working within two R waves. This was an ideal choice, as it was known that either end of this window is certain to be within the active zones by definition.

1. **Filter Signal**

   The absolute value of the signal (with the overall mean of the signal as a zero line) was filtered by averaging, giving an almost sinusoidal form of signal, with but one minimum between the two R peaks. Specifically, with an input signal y(t), each point in the filtered signal f(t) was a sum of all the points within the gate of τ.

$$f(t) = \sum_{i=0}^{\tau} y(t+i) \qquad \textit{Equation 4-1}$$

Typically τ = 300mS.This filtered signal was then scaled between 0 and 1.

2. Find the maximum and the minimum

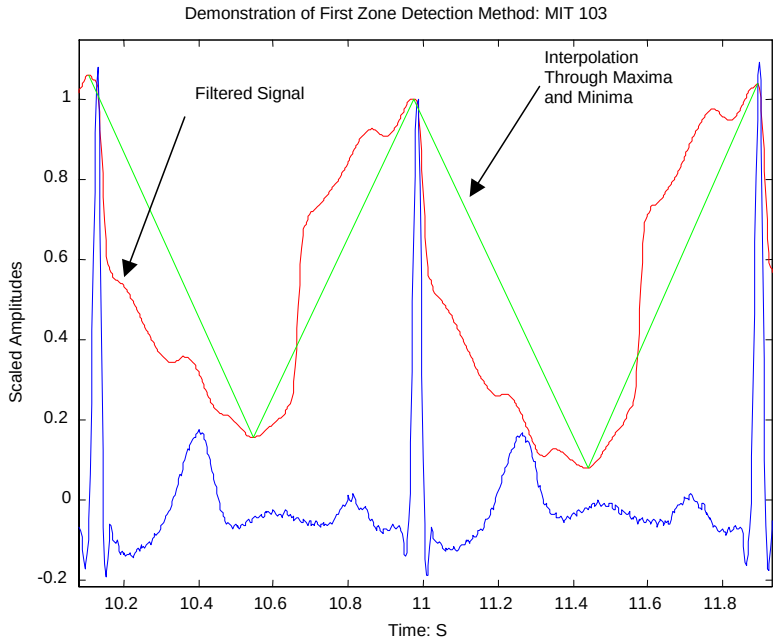

Figure 4-3: *First detection method: demonstration of filtered signal and the linear interpolation between minima and maxima*

The two maxima of each window were then found (corresponding to points near the R waves), and the minimum of the signal (usually corresponding to a point of low activity in the T-P segment. These points were then joined via a linear interpolation.

3. Intersection of interpolation with threshold

The inactive zone is then chosen by cutting the ECG window at the intersection of the interpolation, and a threshold at 30% of the height of the R. (A value of 30% was chosen arbitrarily as this method was never fully optimised)

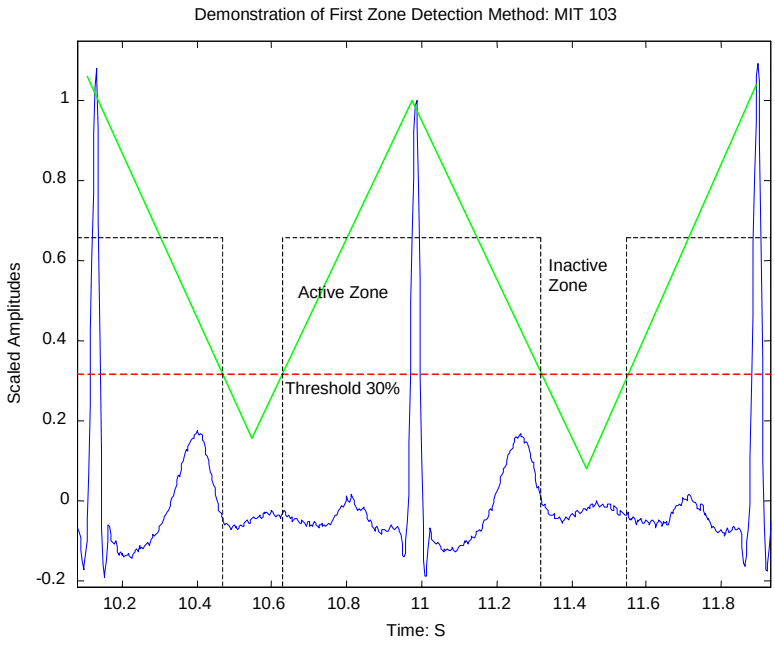

Figure 4-4: *First detection method: demonstration of choice of inactive or active zone via a threshold at 30%*

See figure 4-4 for a graphical representation. This simple approach worked surprisingly well for normal ECG signals though gave unpredictable results with more complicated pathologies and higher noise amplitudes. This inaccuracy was partly due to variations from remaining low frequency noise, causing the threshold to give arbitrary results.

The principle ideas that were gained from this were:

- A need to select the region of noise between T and P waves
- Use of a variable threshold to determine important regions

4.2.2 Direct Zone Method

It was evident from the above method, that a more direct method was needed to take into account all of the possible features in a pathology signal, rather than just assuming that the temporal position of the noise will be around the middle point of two R waves. The algorithm takes four steps:

1. **ECG window split above and below baseline:**
   The signal was first split into two sections above and below a zero value (signals A and B respectively). This zero value was chosen as an estimate of the baseline, and was calculated by taking the mean of the ECG window. This method enables a waveform to be defined by its crossing of the isoelectric line, and so emulates the approach of the cardiologist.

2. **Filtering each section individually:**
   Signals A and B were then filtered separately again using equation 4-1 summing over local points (τ = 16mS), thereby reducing the amount of high frequency noise and hence

the number of regions detected (see criteria 2 above). If the absolute values of the signals had been used in place of splitting the signal, the features of the P wave and the T wave would have become washed out in this filtering method, due to their close proximity to the large valued R waves.

3. Find maxima in each filtered signal:

The zones in signals A and B were then defined by looking for maxima in the signal, defined by zeros in their first derivative (NB. Signal B was inverted).

4. Find end points of each zone:

There were two possible ends for each zone; the start of a region of zeros (caused by the splitting), or a minimum in the signal. Whichever of these occurred first, before and after the maximum, was taken as the end of the zone (see figure 4-5).

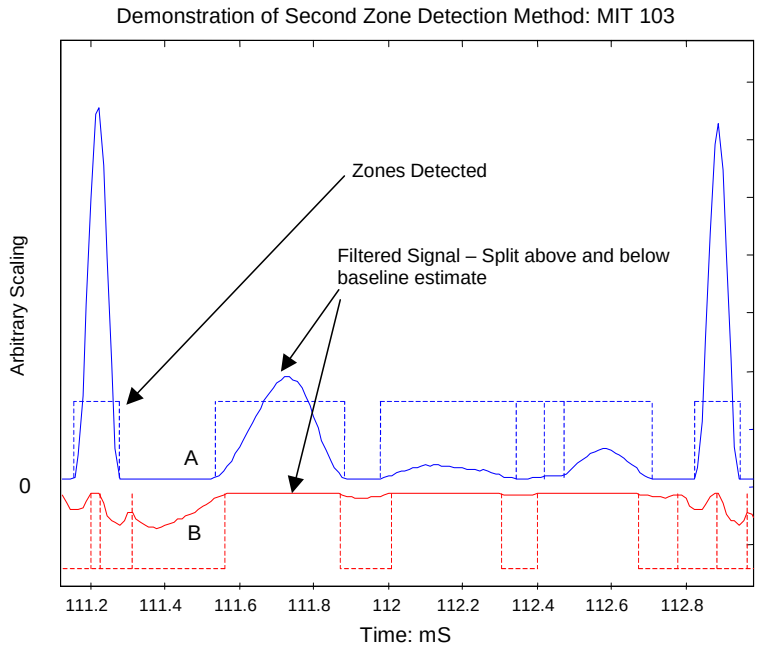

Figure 4-5: *Second detection method: demonstration of ECG signal split above and below a zero, and the subsequently detected zones*

In this method each zone was assigned an importance value via the absolute value of the filtered signal at the positions of the maxima. The activities of these zones were then decided via a varying threshold acting on these importance values, in a manner similar to that which will be described in the next chapter.

This method of zone detection was however limited in numerous ways. The most evident problem with the process was its dependence on a good approximation to the isoelectric line from which to split the signal. In fact, this need led to the author building a baseline detection algorithm and an ancillary algorithm to determine the accuracy of the regions.

In addition, the filtering method was not robust enough to cope with the large variability and noise of the ECG signal, with the optimum number of summation points varying considerably. Hence, it was possible for too few zones to be detected, which enabled inactive zones to be included with active waveforms, thus allowing them to pass the threshold.

However, the most limiting factor of the system was the need to calculate and filter the signal for each chosen window (area between the R waves), thereby increasing the calculation time linearly with the number of baseline calculations required. This could be as many as four times in each window. To keep the computing time considerations within bounds, the rejection criteria had to be relaxed to an unacceptable extent.

The principle beneficial ideas that were gleaned from this method were:

- Pre-filtering the signal to reduce the numbers of zones
- The necessity of looking for noise in all inactive segments, not just in the T-P segment
- Using the derivatives of the signal to subdivide the signal into sensible segments
- Using the maximum excursion from a baseline estimate in each zone to determine the zone's relative importance

4.3 Zero Crossings- Marr-Hildreth Algorithm

The next significant improvements derived from researching the limitations of the previous methods. This breakthrough came from the domain of computer visualisation technology, where the need to detect the edges of shapes in a 2-dimensional image is important. The problem in the active zone case under consideration, is simply the one-dimensional version of this. The algorithm used is that discovered by Marr-Hildreth [Marr and Hildreth (1980)].

4.3.1 Choice of Filter

Within this algorithm, the optimum filter for edge detection is chosen on the basis of two criteria:

- **The filter must be smooth and roughly band limited in the frequency domain**- In our case, this will reduce the number of changes in frequency, and hence number of zones.
- **The filter must be temporally localised** – This for us will mean that the filter at each point is only affected by points in its local vicinity. For example, the filter on a P wave will be a smooth average of adjacent points, and unaffected by the changes in the R wave.

Thus, the filter must be localised in both the time and frequency domains. This implies a form of uncertainty expression between these two conflicting statements. The only filter to optimise this relationship is that of the Gaussian.[Babaud-J, et al. (1986)]

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-x^2/2\sigma^2\right) \qquad \textit{Equation 4-2}$$

The Gaussian g(x), is then convolved with the signal y(x):

$$f(x) = g(x) \otimes y(x)$$
$$f(x) = \int_{-\infty}^{+\infty} y(x)\, g(x-\tau) d\tau$$

Equation 4-3

The only variable parameter is σ which effectively widens the Gaussian, hence producing a sum over more points, reducing its localisation, and smoothing out the features in the filtered signal. As can be seen in figure 4-6 the fit of the filter to the true signal is extremely good, and it is able to model through the noise. There is however distortion around the higher frequencies in the QRS complex, but this is not a problem as it is only the points of inflexion that are important in defining the zones, not an accurate model. The model also picks out all of the important peaks and troughs of the signal, giving an exact correlation between their position in time on the ECG signal, and that in the filtered output signal. This is vital when determining the zone importance numbers (see section 4.4).

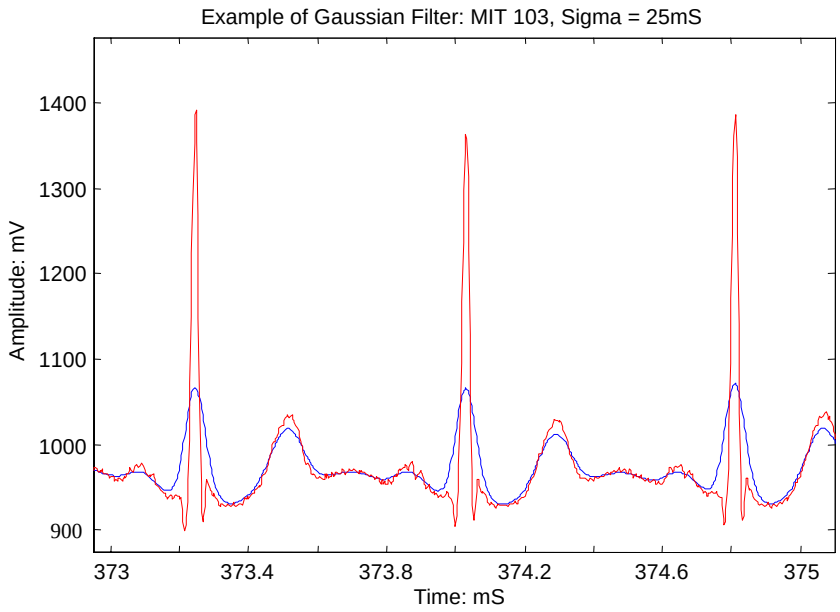

Figure 4-6: *Demonstration of results from Gaussian filter with a σ of 25mS*

4.3.2 Cutting points

Once that the filtered signal is established, it must be cut into sensible sections (region criteria 2). This is achieved by considering the zero-crossings of the second derivative of the signal, ensuring that the waveforms are never cut into separate regions.

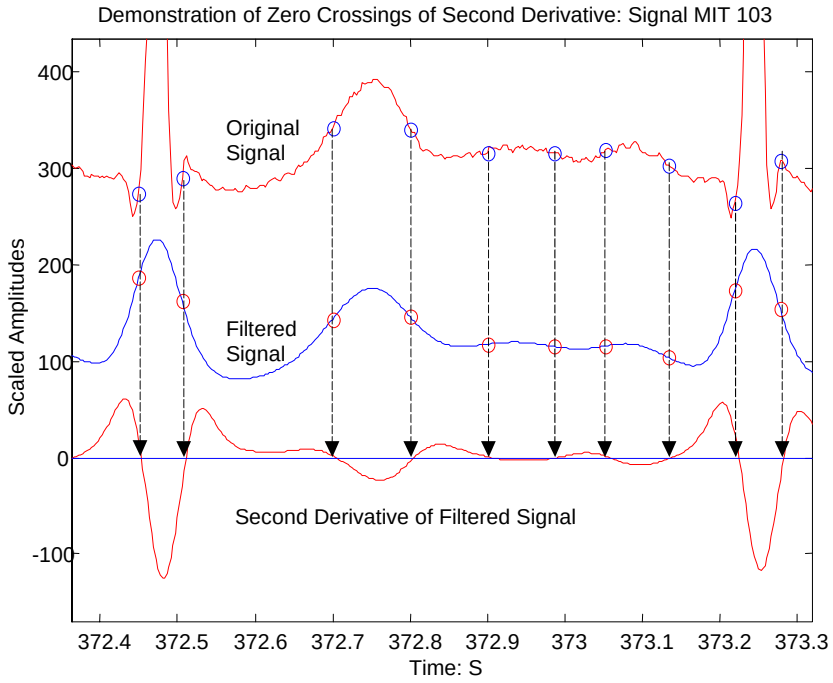

Figure 4-7: *Demonstration of zero crossings of second derivative, showing the accurate prediction of meaningful inflection points in the unfiltered ECG signal.*

4.3.3 Choice of σ

With the previously used filters, there was a great deal of variability in the quality of the zones detected (with reference to the criteria above). Hence, as the scale of the filter is increased (increasing τ), the number of zones decreases almost linearly, hence demonstrating no optimum choice for σ. This is not true of the Gaussian filter, as can be seen in figure 4-8, a graph of the relationship between the average number of zones detected per window, and the choice of σ for the entire MIT database. There is an evident period of stability between 19mS and 35mS, where all the traces remain at around 7-12 zones per window. This corresponds to the figure desired in criteria 1 from the beginning of this chapter.

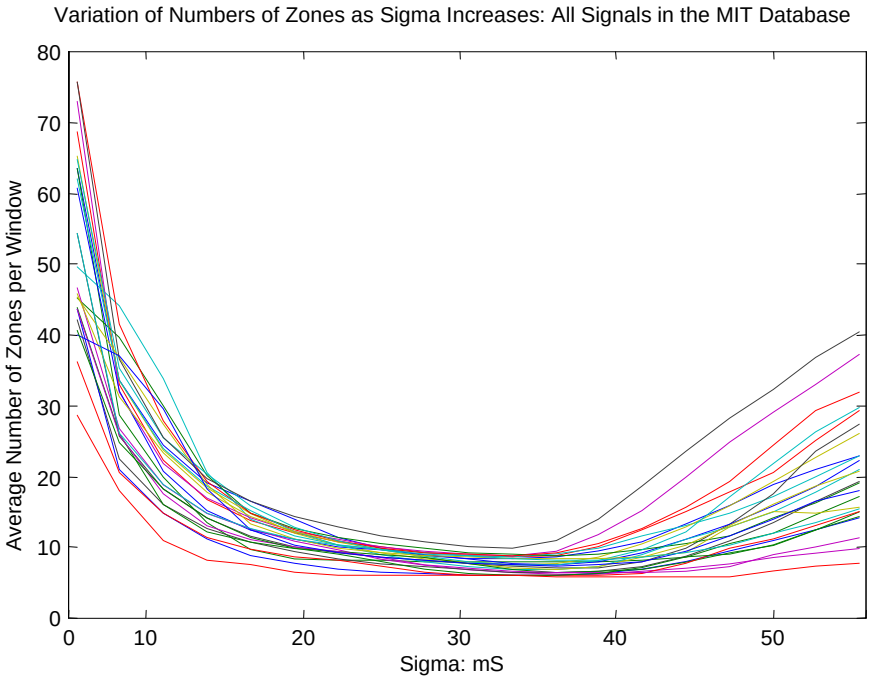

Figure 4-8: *Graph of average number of zones per ECG window for all signals in the MIT database, demonstrating a region of stability after 20mS, and an unexpected rise after 40mS*

As is seen in the graph, there is a rise as the value of σ increases. This is an unexpected result and is interpreted as an increase in points of inflections without corresponding maxima (see section 4.4.1). As such, this corresponds not to an increase in the number of features identified per window, rather the over smoothing of the ECG. As such, we are seeking a value that is before this rise.

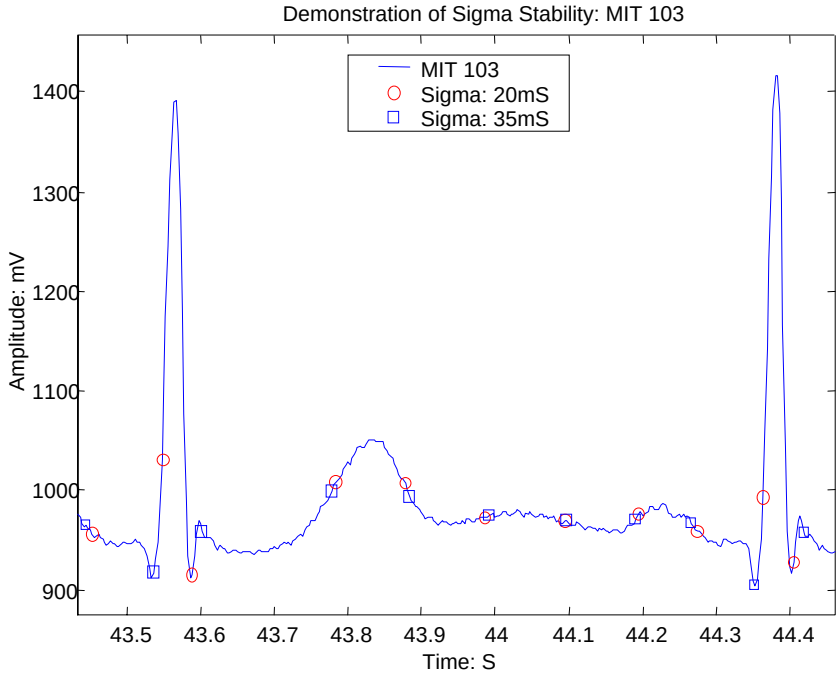

Figure 4-9: *Demonstration of the lack of variation in choice of inflection points, given two choices of sigma at either ends of the identified stability zone in figure 4-8.*

Using this stable zone, the robustness of the σ choice can be shown. As is seen in figure 4-9, with σs chosen at either ends of this zone, the position of the cutting points hardly varies at all.

Using this information, and the results from figure 4-8, a σ of 25mS was chosen. This value can then be applied to filters at whatever sampling frequency.

4.4 Importance Value

Once these zones are chosen, a value is assigned to each to represent its activity level. These amplitude values are defined at the zero crossing of the first derivative of the filtered signal located on the time scale (see figure 4-10). The actual importance value given to the zones, are precisely located at the corresponding points in time on the actual ECG signal, thus eliminating the error seen at high frequencies in the Gaussian filtered trace (see figure 4-8).

N.B. At this stage of the algorithm, the importance values are defined from an arbitrary zero. These values will be later scaled and centralised (choice of an apt reference zero) in the baseline detection algorithm.

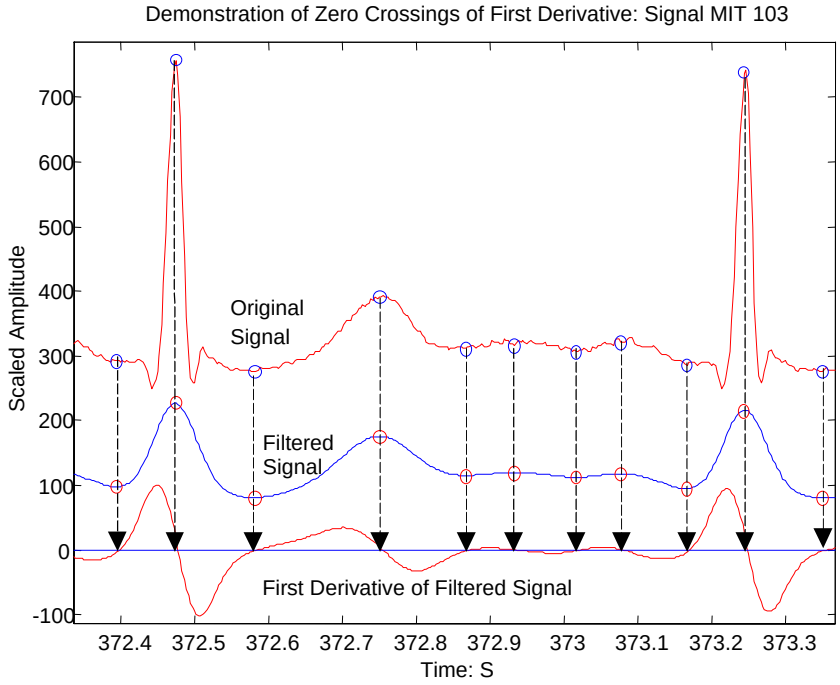

Figure 4-10: *Demonstration of how the importance values are chosen, using the maxima and minima of the filtered signal.*

4.4.1 Points of inflection with no maximum

There are cases in the signal, where there are points of inflection, without a corresponding maximum between them. See figure. In this case, for reasons of time in the algorithm, a maximum is created in the same position as the troublesome inflection point and that of the next inflection point (shaded points in figure). It is much faster to use the known inflection points, rather than take each problematic situation and look for the best possible point to represent the zone. These are then able to satisfy the definition of a zone: a maximum/minimum, bordered by two inflection points.

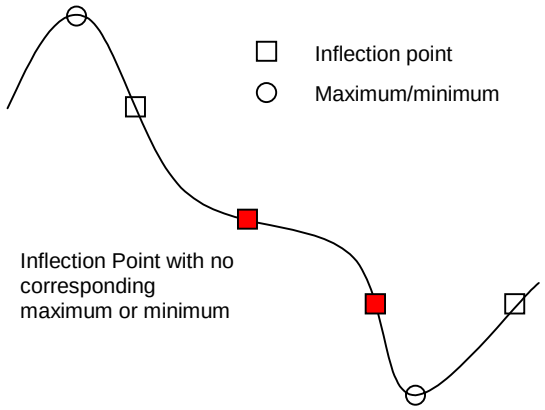

Figure 4-11: *Figure demonstrating the occurrence of inflection points without corresponding maxima*

It is possible for many of these situations to occur in the signal, depending on the filter's interpretation of the inactive zones, where the trace is effectively flat.

The next chapter, the baseline algorithm, discusses the way in which a decision is reached over which of these zones contain waveforms, and those that correspond to the isoelectric line.

4.5 Results

This section demonstrate two principle results: the ability of the filter to distinguish the importance waveform variations beneath high and low frequency noise, and the errors induced by the treatment of inflections without maxima as entirely new zones. Results have been chosen from the MIT database and from the ELA test signal, to represent qualitatively the accuracy of this process. In each graph, the zones are distinguished by a square wave.

4.5.1 MIT Database

Signals were chosen so as to represent the various difficulties that are presented in baseline detection.

Normal Signal

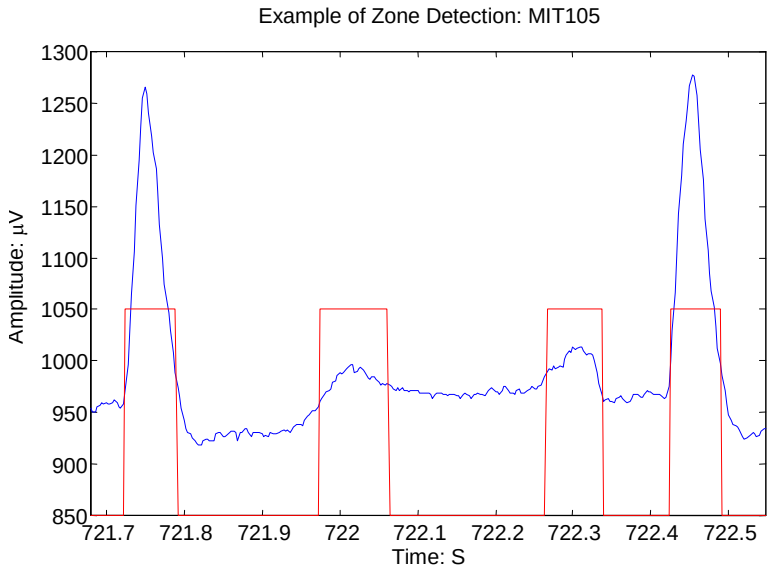

Figure 4-12: *Accurate zone detection for a normal ECG signal*

Figure 4-12 demonstrates correct zone detection for a relatively normal signal. The signal is separated into an R wave, ST segment, T wave, TP segment, P wave, PR and the final R wave, with no extra waves.

High Frequency Noise

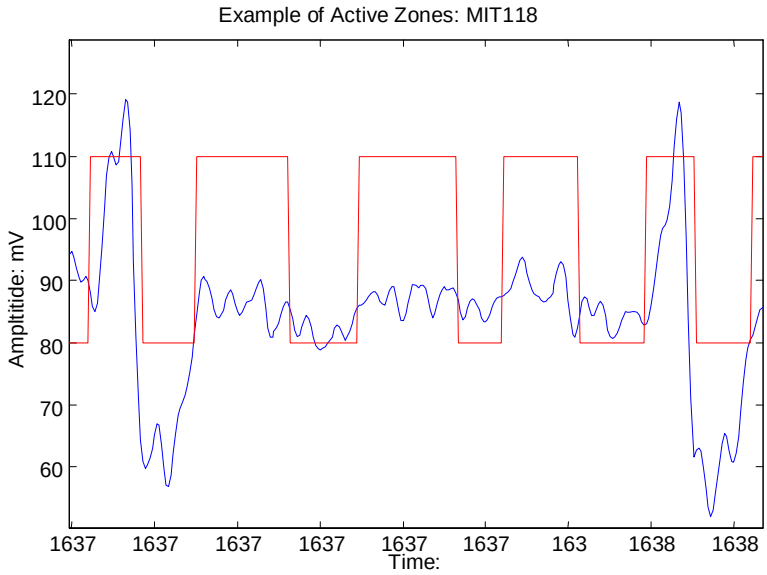

Figure 4-13: *Correct detection despite high amplitude noise*

Figure 4-13 demonstrates the ability of the filter to distinguish high amplitude noise from interesting waveforms.

Low Frequency Noise

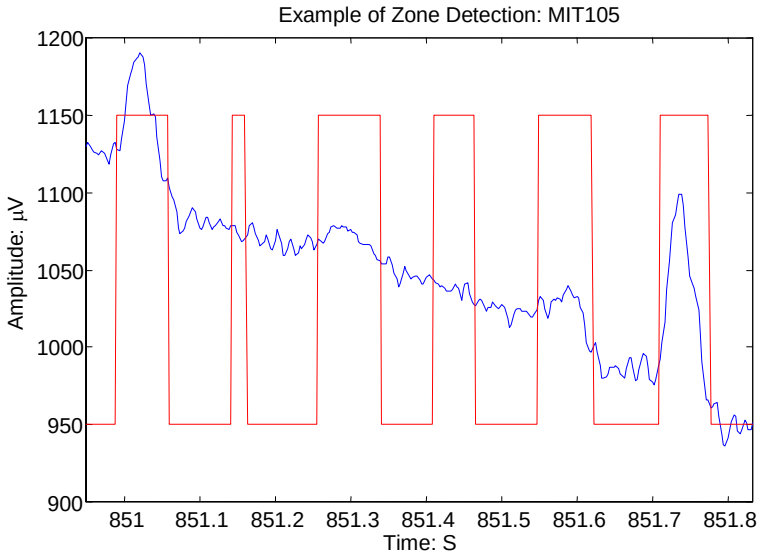

Figure 4-14 *Correct detection despite low frequency noise*

This next case demonstrates the filter working despite the slope given by variations in the low frequency noise. There are two additional zones, but these are caused by inflections without maxima, owing to the gradual decreasing of the filter, rather than a falsely found waveform.

4.5.2 Widely Varying Waveform Size

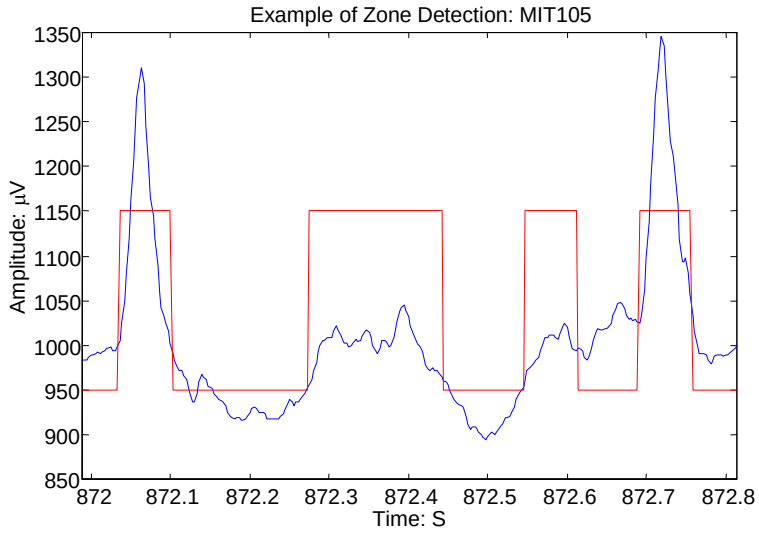

Figure 4-15: *Correct detection despite large variations in waveform size*

Again, in figure 4-15, the filter has correctly not given zones to the small waveforms present due to noise on the larger waveforms. This reduces the risk of separating a waveform into several zones.

Inflection points without Maxima

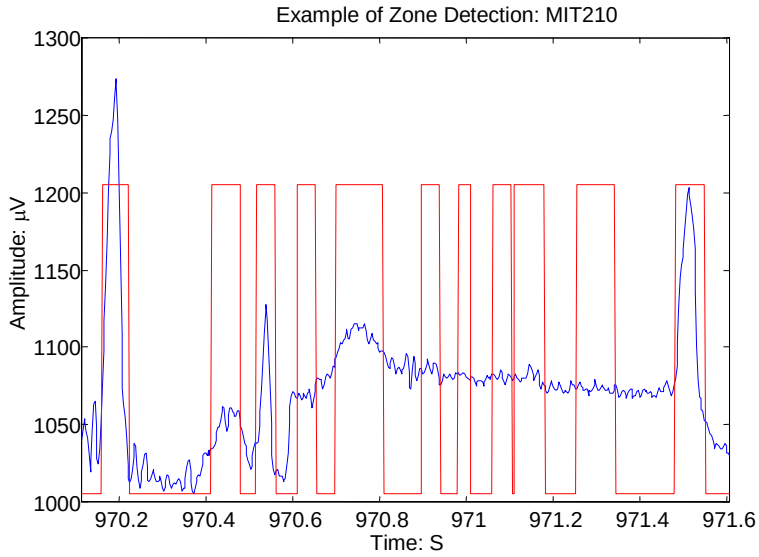

Figure 4-16 *Demonstration of the large number of zones added due to inflection points without maximums*

This last example from the MIT database (figure 4-16) demonstrates the difficulties given by the phenomenon described in section 4.4.1. In this area of gradual gradient decrease, there are many inflection points, but no maxima. This leads to the signal being cut too many times. Fortunately, this primarily occurs in the flatter periods of the isoelectric line, and so there is a low risk of splitting waveforms. The fault does however increase calculation time.

4.5.3 Results from the ELA Test Signal

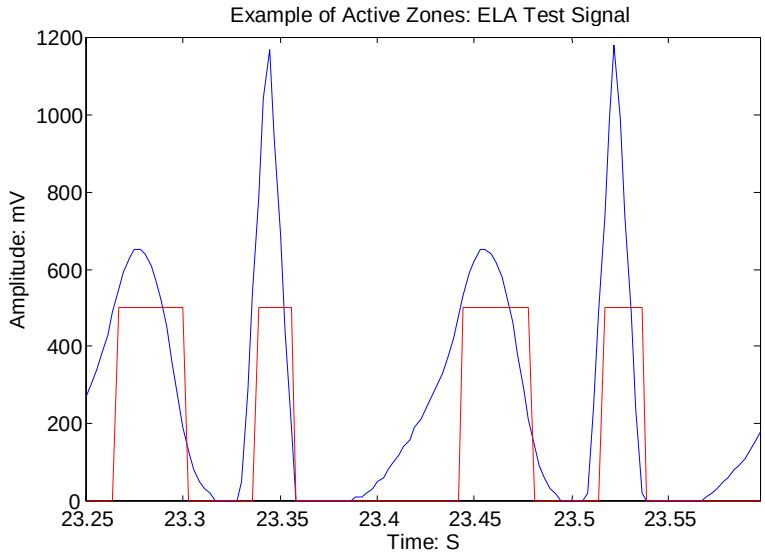

Figure 4-17: *Incorrect zone start and end points in ELA test signal, due to very high frequency heartbeat*

The signal from ELA is an artificial signal and presents a difficult problem, i.e. the rapid heartbeat of a very young infant. As can be seen in this example (figure 4-17), the choice of inflection points to cut the signal is inadequate. Nevertheless, all of the inactive zones will have the correct importance value, allowing each zone to be correctly identified as inactive or

active, though the zone will contain many high amplitude values due to the waveform being split incorrectly. This leads to outliers in the baseline detection algorithm, and the possibility of errors. This error lead to numerous changes in both the noise estimation and baseline algorithms, notably the use of the median instead of the mean, to calculate a baseline estimate, and has thus been resolved.

4.6 Summary and Conclusions

The important points derived from this analysis are:

Method

- Active zones can be used to enable signal treatment that does not affect the waveforms
- The best results are obtained if the ECG is filtered with a high pass filter prior to zone detection
- A Gaussian filter with a σ at 25mS provides the optimal filter for zone detection
- The zones are defined by the zero crossings of the second derivative of this filter
- The activity levels of each of the zones are defined by the zero crossings of the first derivative

Results

- Very accurate zone detection in difficult situations (pathology type, high frequency noise, low frequency noise)
- Very robust choice of σ
- Ongoing difficulties with inflection points without maximums

In conclusion, the use of the zero crossings to cut the ECG signal into zones has been shown to be extremely accurate and robust when dealing with the incredible variations that are possible from heartbeat to heartbeat.

4.7 Improvements and Further Study

There are several improvements and areas for further investigation that could be made to this system in light of the summary and conclusions above. These have not been achieved within this project due to time constraints.

- **Independent of baseline**

The dependence on a correct first estimate of the baseline is the principal drawback of the baseline algorithm proposed in the next chapter. Whilst this will become clearer in the next chapter, it is within this section that a solution may lie. The issue lies with the estimate of the 'importance values', as an estimate of the baseline is required to give an unbiased scaling (see section 5.4). A perhaps better method of calculating these values would be to take greater account of information within the zones such as its relative duration. Another method may be to look at the slope of the line from the current zone, to the neighbouring zones, thereby

looking at the frequency of the zone as a discriminate. For example, the QRS complex will have a very small width and high amplitude, and hence give a high frequency, whilst noise in the TP region will have a relatively large width, and low amplitude relative to the next zones. These parameters of width and relative amplitude could then be used as the threshold parameters in the baseline algorithm.

- **Study of optimal σ**

A fuller study of the choice of σ should be made, taking into account the appearance of an unexpected minimum, and the full effects of differing sampling frequencies on the optimal choice of σ, should be studied. In addition a study of how using discrete time samples, and hence a non-smooth Gaussian curve at small values of σ, affects the quality of the filtered signal.

- **Interpretation of number of zones**

Another study may include information on how the number of points of inflection, and hence number of zones gives an indication of the quality of the zone approximation. For example if there are too many zones in the signal, is this an indication of too much noise.

- **Points of Inflection with no Maximum**

As discussed in section 4.4.1, inflection points corresponding to a smooth ascending/descending curve do not have a maximum. Currently, to keep time costs down, a maximum is created to account for this, though this leads to small zones, and inaccurate importance values. These are of little use to the zone selection process. A better method would be to take into account the information within the zone, such as its width, and relative amplitudes, and therefore either make it into a zone, or join it onto a neighbouring zone. In addition these inflection points may provide some indication of isoelectric line placement, as it is usually in the low amplitude areas of segments, that this phenomenon occurs.

Chapter 5. Baseline Detection

This chapter aims to calculate a baseline that will take into account the baseline wandering of the signal, and give an isoelectric line from which the wavelet analysis can take place. This will be achieved using the zones detected previously to determine inactive areas of the ECG signal, which will be correct in spite of any baseline wandering.

5.1 Importance mathematically

The importance of the isoelectric line is linked to the manner in which the cardiologist measures the amplitude and duration of the ECG waveforms. The cardiologist will simply place a ruler across the signal, and make a best estimate using historic experience, and taking into account factors such as the waveform shapes and their relative positions, as well as the amount of baseline wandering present. He will then measure the amplitude of each of the waves from this baseline estimate, and be able to define each of the different waveforms present in the signal.

For an automated system, this brings a vicious circle of needing the pathology to calculate the baseline, but needing the baseline to interpret the pathology. In most automated systems, this problem is removed by taking a fixed point, a certain distance behind the R wave. This corresponds to a position within the PR segment. The algorithm presented here aims to take into account factors similar to those of the cardiologist.

In the system proposed by the research group, the isoelectric line is of vital importance for the correct modelling and classification of the waveforms. This is shown in two ways: from the geometrical interpretation of wavelet analysis and an example of poor wavelet selection with an incorrect baseline.

5.1.1 Geometric Interpretation

By returning to the geometrical interpretation of wavelet analysis, as developed in chapter 3, the decision on which is the most pertinent wavelet to model the signal, is shown to be dependant on the choice of the baseline. Again we begin by representing the ECG signal window by a vector **Y**, and two wavelets by $\mathbf{W}_1$ and $\mathbf{W}_2$. In addition we are also able to define a new vector in the same dimensional space, that of the baseline vector, **B**.

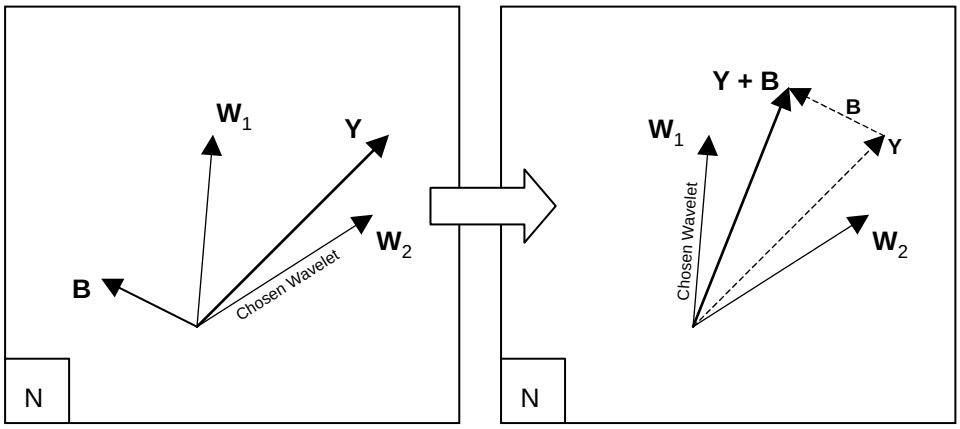

Figure 5-1: *Geometrical demonstration of Gram-Schmidt dependence on correct baseline detection*

We first assume there is one correct choice of wavelets from the library. As is seen in the diagram, the addition of **B** to the signal **Y** has caused the wavelet $\mathbf{W}_2$ to no longer be the chosen wavelet. Thus it is demonstrated that the choice of **B**, affects the order in which the wavelets are chosen. This proves the dependence of the wavelet selection on choice of baseline.

5.1.2 Experimental Example

This dependence can also be shown with an example from the modelling system currently used by the research group.

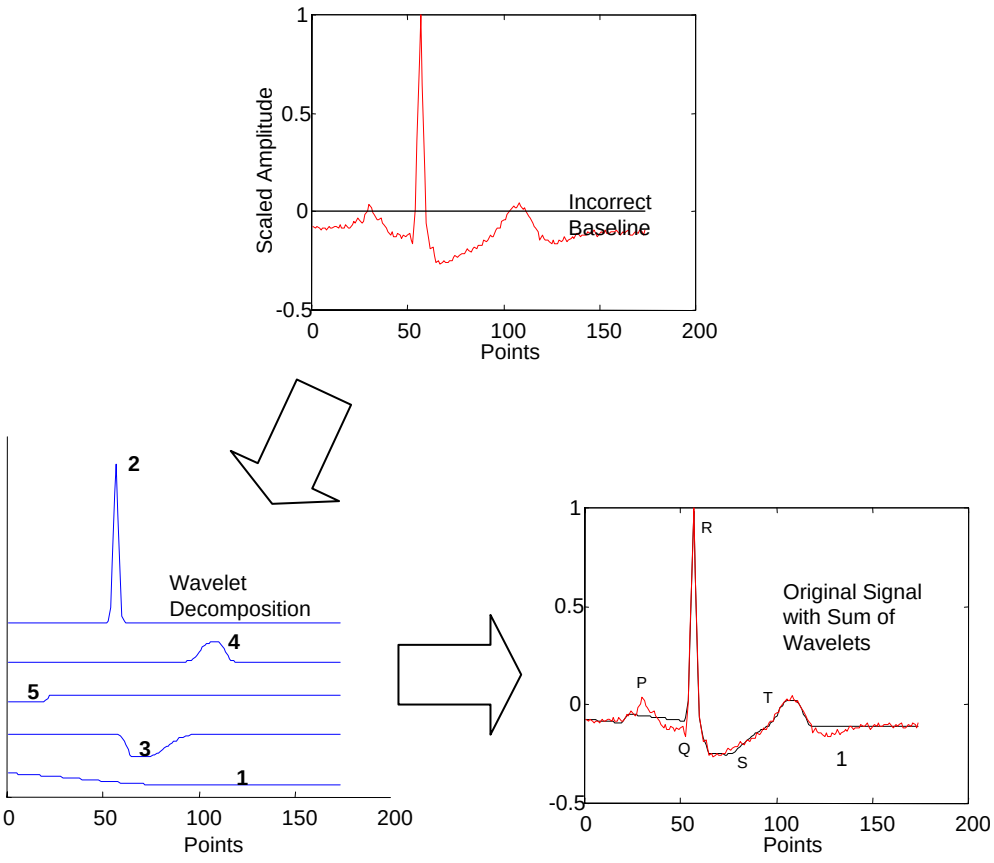

Figure 5-2: *Demonstration of poor wavelet selection with an incorrect choice of baseline*

It is seen in figure 5-2, with an incorrect isoelectric line choice, the wavelets chosen are not representative of the waveforms contained in the ECG signal. For example wavelets 5 and 1, have been selected from the library, rather than wavelets to model the Q wave, or the feature identified in the lower right figure by 1.

N.B. In terms of project progression, this algorithm of baseline detection was introduced initially to deal with errors in zone determination in section 4.2.1 (Triangle Fit of Maxiums), with the aim of creating accurate zones for the algorithm discussed in appendix A1 (Wavelet Filtering). This was before the need for an accurate isoelectric line was made evident.

5.2 Difficulties

The ECG signal is of an inherently difficult form for accurate analysis. Some of these reasons have been outlined previously in the discussion of active zones, however there are several which are specific to baseline detection.

5.2.1 Pathology Shape

The most evident problem is the wide variability of pathology shape, hence one particular set of criteria for discerning the baseline may not be valid for another. Thus the parameters chosen should be universally applicable, and a way of determining the quality/accuracy of the calculated baseline must be found.

A more specific difficulty is the absence of an obvious period of inactivity in the ECG window (see figures 2-4 and 2-6). Examples of this can be found even in patients with a normal heartbeat and results from high frequency noise distortion of the flatter sections of the signal. In more complicated pathologies, there may be no distinguishable period of pause in the depolarisation and repolarisation of the heart. The extreme case is that of fibrillation (figure 2-7), where there is no evidence of the isoelectric line and the mean of the window is taken as the measurement.

5.2.2 Low frequency noise

The large amount of low frequency noise usually present and the edge effects of the high pass filter, cause their own distinct problems. It becomes possible in numerous cases that a period of inactivity can be construed as an area of importance due to the inclination it is given by low noise (see figure 5-3). In addition, owing to the nature of the high pass filter, a flat section in a normal signal can be given a non-zero gradient after all the low frequencies have been removed. This filtering problem is notable in artificial signals such as the ELA test signal.

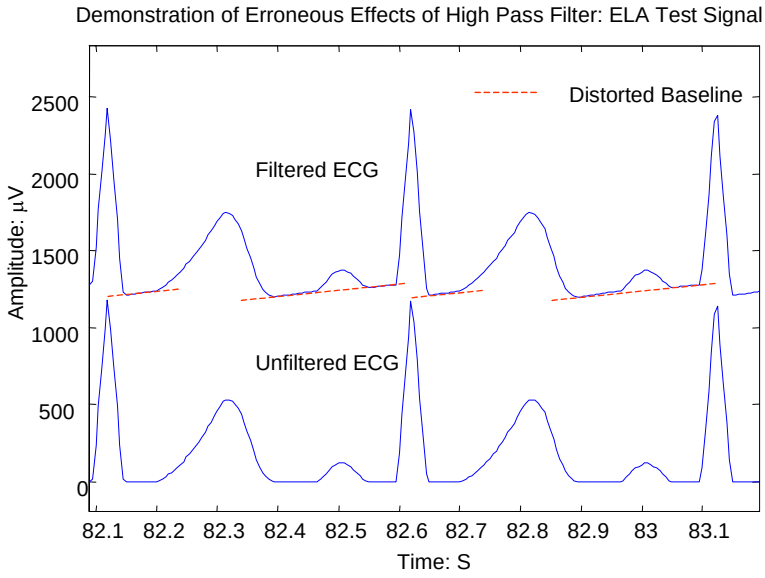

Figure 5-3: *Distortion of artificial signal due to high pass filter*

5.3 Previously Used Methods – Baseline Wandering

A distinction should be made here between removing baseline wandering, and the detection of the baseline. Baseline wandering is low frequency noise, present in the signal due to motion of the heart axis (or problems with the electrode), causing frequencies in the region of 0 to 5 Hz. Whilst the author has found many references that deal with the removal of baseline wandering, none have been found to directly calculate the isoelectric line. The field of baseline wandering is very important in the manufacture of electrocardiograph machines, as well as in other computer-automated systems. This section will review other already tested methods for baseline removal, and their success and disadvantages. A comparative study has been made in [Royo (1998)]

There are three distinct methods currently used to remove the influence of low frequency interference:

5.3.1 Cubic Spline

The first method, proposed in [Meyer and Keiser (1977)], uses a polynomial interpolation to join up points chosen from the PR segment. The baseline is estimated via the cubic spline polynomial, which uses information from the current window, and that of the precedent. This then subtracted from the original signal, thus reducing the signals baseline wandering. Flaws in this system come from the choice of points in the PR segment. This is discussed further later on in this chapter (see section 5.3.4). Improvements to this method are discussed in [Froning, et al. (1988)]

5.3.2 Digital Filtering

It is known the majority of the interference will have frequencies in the range 0-5Hz. It is possible to remove these using a standard high pass filter, however this risks a deformation of the signal due to an overlap of the frequency of an interesting waveform with that of the

noise. Having a cut off frequency of 0.5Hz reduces this chance, though deformation of the ST segments is still seen. [Shusterman, et al. (2000)]

An improved method proposed by [Sornmo (1993)], involves a time varying linear digital filter. In this case the filter is varied from heartbeat to heartbeat window, with the cut off frequency being controlled by the amount of baseline wandering within that window. It was found to produce better results than those of a standard linear filter, as well as that of the cubic spline as described before.

5.3.3 Wavelet Analysis

More recently, has been the use of wavelets in the domain of baseline wandering. The wavelets decompose the ECG signal into 7 frequency bands using a wavelet. The last of these bands, containing the lowest frequencies, is then used a reference signal in an adaptive filter. This produced better results than both the standard and adaptive filters described previously, though is time consuming. [Li, et al. (1995)]

5.4 Previously Used Methods –Isoelectric Line detection

PR segment analysis is the most commonly used method for estimating the isoelectric line in automated Holter systems, with a point at a fixed distance before the R wave is chosen. The distance usually chosen is 40ms before the R and corresponds to a point within the P-R segment. The major advantage of this system is the ease of calculation with no need for an algorithm; 40ms is simply removed from the already calculated R-wave positions. The disadvantages are numerous however.

- **The system relies on perfect R wave detection**

In certain circumstances, the position of the R wave can be mis-detected. For example in the case of an overly large R wave (e.g. PVC see figure 2-5), the R wave is usually detected several ms after the actual value. As such, the point 40mS behind this, falls onto the slope of the R. This in effect, raises the baseline estimation over the size of the P and T waves, and the wavelets detected will be modelling the inverse of the signal.

- **No discernible period of inactivity**

Due to the wide variability of ECG signals, there is often no period of calm in the hearts depolarisation and repolarisation, and it becomes likely that the point chosen will fall in the trough of the Q wave, or even on the P. This problem is evident in cases such as fibrillation where there are no waveforms, or segments. In this case 40mS before the R wave will give a random baseline.

- **PR segment distorted by noise**

If there is a large amount of noise in the signal, the difference between the point falling in a trough and a point falling on a peak can be extreme. The result is a baseline that is not consistent.

This method of isoelectric line detection was tested against the algorithm proposed by the author, the results of which can be found in appendix A2.2. These results represent examples of all of the difficulties outlined above.

5.5 Proposed Base Detection Algorithm

The principle aim of the algorithm is to apply intelligent thresholding to the zones and importance values found in the previous chapter. This corresponds to determining the regions of the signal that represent an active waveform (PQRSTU) and vice versa. Depending on the specific signal, the inactive zones will include the P-R, S-T and T-P segments. By interpreting all of these regions and not just that of the P-R (as described earlier in chapter 4), the best estimate of the isoelectric line can be calculated. Later, these inactive zones can be used to estimate the amounts of high and low level frequency noise in the unfiltered ECG signal (see chapter 6). An additional more practical aim is to keep the calculation time down to a minimum.

5.5.1 Summary of Algorithm

As can be seen in figure 5-4, the algorithm can be split into three distinct modules. The passage from module to module is determined by the accuracy with which the zones are found in the current module. For example, in a perfect signal the algorithm will remain entirely within module 1, where there are limited calculations needed, and hence keep the calculation down to a minimum. If however the initial module does not pass the accuracy percentage, it will pass to module 2 and maybe to module 3 if extremely difficult, where more adapting calculations take place. The decision process is best explained by referring to the different parts of the flow diagram.(Markers 1-5 in figure 5-4)

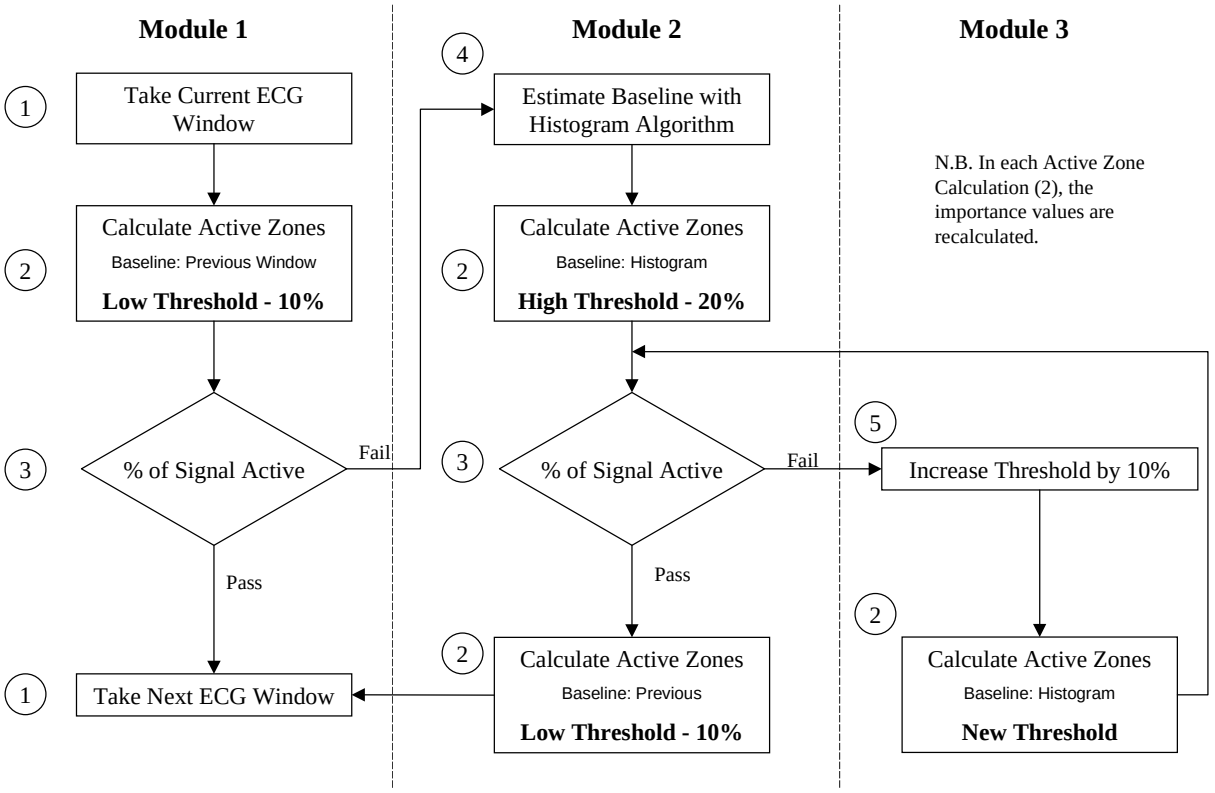

Figure 5-4: *Flow chart representation of baseline detection algorithm*

5.5.2 Explanation of Algorithm via Flow Diagram

The flow diagram can be split into five distinct calculations (referred to in figure 5-4).

1. Choice of ECG Window

The initial step in the algorithm is to segment the ECG signal into manageable sections or 'windows'. These are chosen intelligently so that if the baseline jumps considerably between heartbeats, there is minimal deformation of analysable waveforms. With this in mind, the start and ends of each window are the positions of the R waves on the time axis (calculated earlier using standard digital filtering). Thus window (i) is defined as all the points between R(i) and R(i+1), where i = 1, 2, ….N-1, and N is the number of heartbeats in the ECG signal.

2. **Calculate Active Zones**

In this section, the zones are chosen by virtue of whether their importance value is greater than a particular chosen threshold. As is demonstrated in figure 5-5, if the importance values remain as they are from the active zone calculation, there is a bias given to zones that fall upon large amplitude low frequency noise. In the figure, the importance value of a zone representing noise (the TP segment) is shown with higher amplitude and thus greater importance value than that of a P wave (due to the choice of an arbitrary zero as a reference point). Thus, before effective thresholding of the zones can take place, the importance values must be scaled and centralised.

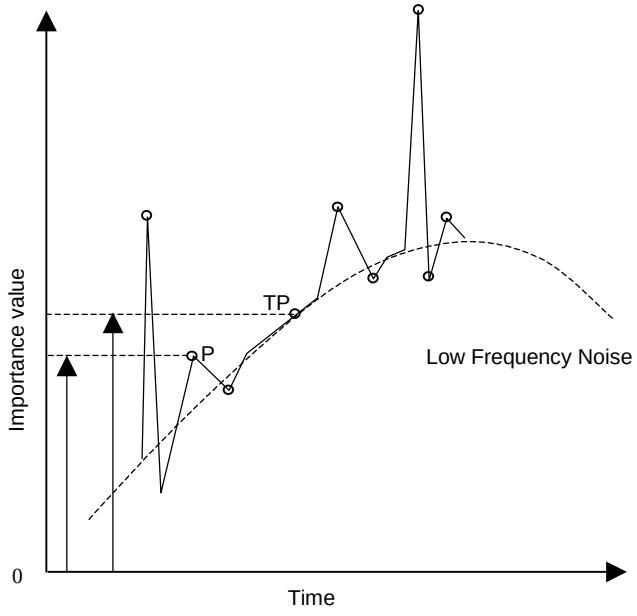

Figure 5-5: *Demonstration of need for scaled importance values, due to the influence of low frequency noise*

To standardise these values, a more pertinent choice of zero is made; this is taken from a first estimate of the baseline. In module one, this estimate is the calculated baseline of the previous

window. This assumes that there is limited baseline wandering, which is predominantly true due to the straightening effect of the low pass filter (see section 4.1.3). In module two however, it is decided that this first estimate is wrong, and an estimate of the baseline is made from the current window using histograms (see step 4 in this description of the algorithm).

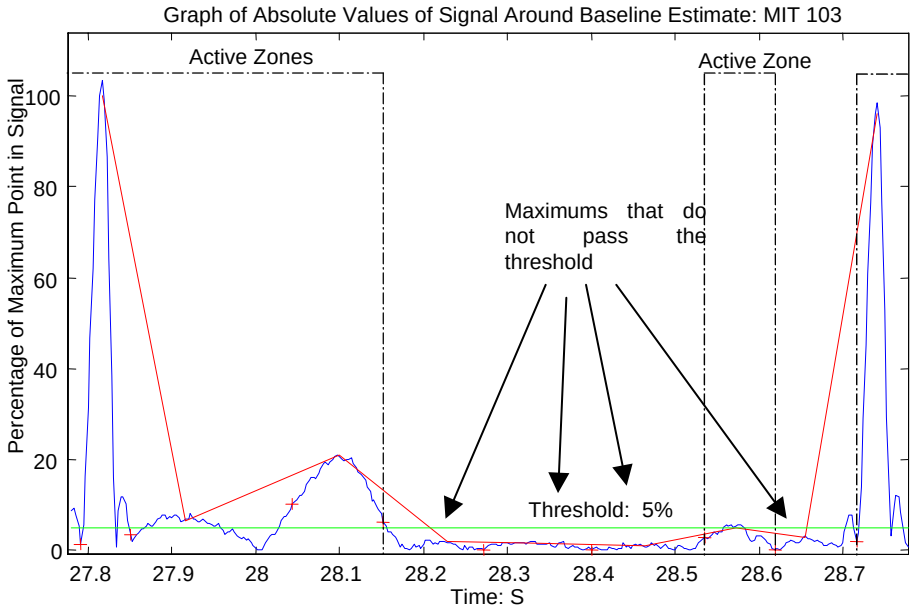

Figure 5-6: *Demonstration of determining zone activity using a threshold, acting on scaled importance values*

The importance values are then transformed into percentages of the maximum excursion point in the window, which is usually an R wave. To reduce the distortion effects of low frequency noise or an overly large waveform (such as the T wave), the average is taken of the current maximum and those of the three preceding windows. The absolute values of these scaled numbers are then taken, so that maximums and minima are treated equally by the algorithm. In summary:

$$S.I.V(i,j) = \text{abs}\left[(IV(i,j) - \text{base estimate}(i)) / \text{mean}(\max(\text{window}(i-3:i))\right] \qquad \textit{Equation 5-1}$$

Where: S.I.V(i,j) – Scaled Importance Values in zone j, of heartbeat i;
I.V. Importance Values (original values from chapter 4)
j = 1,2,…n(i) and n(i) = number of zones in window(i)
i = 1,2… N-1, and N = number of heartbeats.

The activity of a zone is then determined by whether the scaled importance values, pass above a certain threshold (see figure 5-6). This threshold varies from module to module, and is interpreted as a percentage of the size of the R waves.

Once the inactive zones have been selected, the baseline must be calculated. This is done so as allow correct scaling factors for the next window, or next module. For this we treat all of

the points within the inactive zones as an ensemble. It was found that the median of this sample of inactive points, gave the best approximation to the true isoelectric line, rather than the mean. This is due to the zones being cut off at the inflection points of the signal, hence a certain amount of the next waveform will be present in the inactive waveform. Thus the mean of such a zone will be raised due to these data outliers (see figure 5-7).

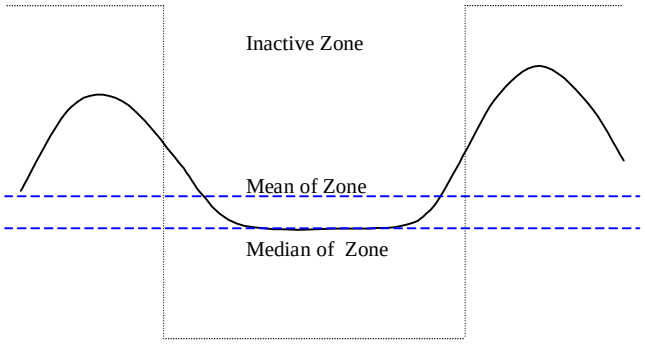

Figure 5-7: *Graphical description of reasoning for choice of median over mean in baseline estimation*

3. Interpret Accuracy of Zones Found

The decision whether to move on to the next window, or to the next module, is made on the percentage of the window that is deemed active. Hence, if over 80% of the window is active, the algorithm will not accept the baseline measurement, and will look for a better approximation. There are two different causes for this failure. The first is a poor first estimate of the baseline, used to scale the importance values. As shown in figure 5-8, nearly the entire signal has been made active, with only the maximum of the T wave not passing the threshold of inactivity. This failure case gives an indication of the quality of the first baseline estimate that was used to scale the importance values.

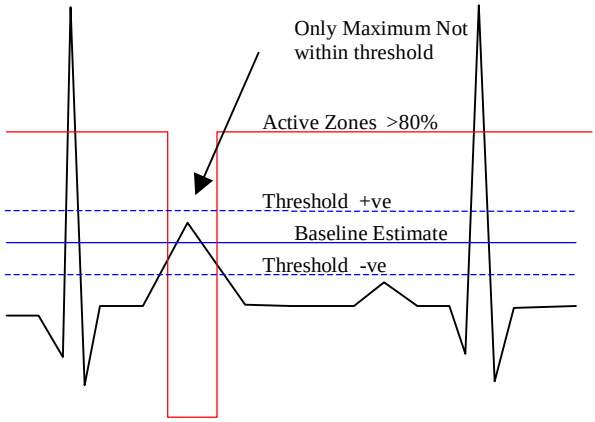

Figure 5-8: *Failure Case 1: Incorrect initial baseline estimate causing incorrect scaled importance values*

Even with accurate baseline estimation, it is possible for the zones to fail the test if there is a very limited amount of isoelectric line visible in the window. For example, in figure 5-9, only the ST segment does not pass the thresholds. Whilst the isoelectric line may be correctly defined from this, the accuracy will be limited due to the paucity of points available for baseline analysis, and noise estimation later on.

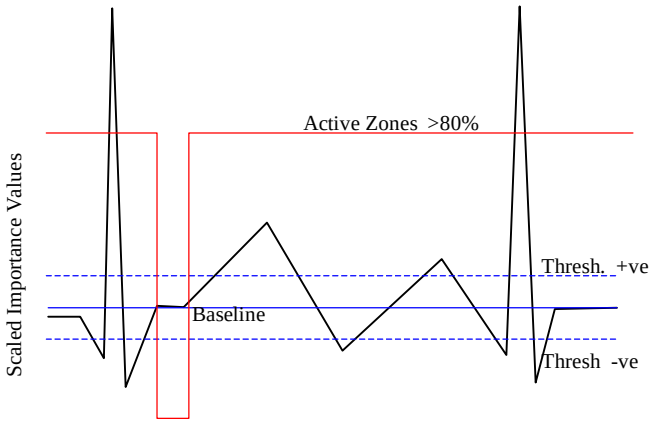

Figure 5-9: *Failure case 2: Too few regions of isoelectric activity for accurate baseline determination*

In this latter failure case, the algorithm works on the hypothesis that a much better calculation of the baseline will be made with greater number of points, even if this means for example that a P and a Q wave are included in the inactive zones for example.

These two failure cases give rise to two needs: a need for a variable threshold to cope with windows where there is limited visible isoelectric line, and a need for a better approximation to the baseline if the first module fails.

4. Estimate Baseline using Histograms

This algorithm is the fail safe of the baseline detection system. It is here that allowances are made for different variations in pathology and noise levels. The working hypothesis is that if the signal is split vertically into several equal sized bins, the isoelectric line will correspond to the bin with the highest frequency owing to the zero gradient in these areas (no depolarisation or repolarisation taking place).

The algorithm commences by taking a wider local view of the window in question, including the current and the precedent windows. This allows a more meaningful statistical sample, and provides backup in the case of sudden jumps in the baseline or flat periods in the signal, which do not correspond to areas of inactivity (for example a flat section on the top of a fibrillation wave).

There are three different outcomes identified in the histogram:

1. **A clear highest frequency (figure 5-10):**

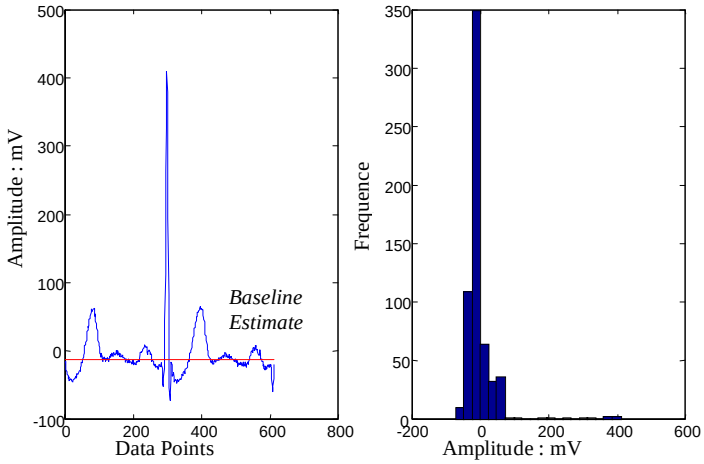

Figure 5-10: *Clear highest frequency*

This corresponds to a bin that is 40% larger than the second largest frequency, and is generally the case in stable normal ECG signals. In this case this bin value is taken as the baseline, and is used as the estimate.

2. An ambiguous highest frequency with a comparable next closest (figure 5-11):

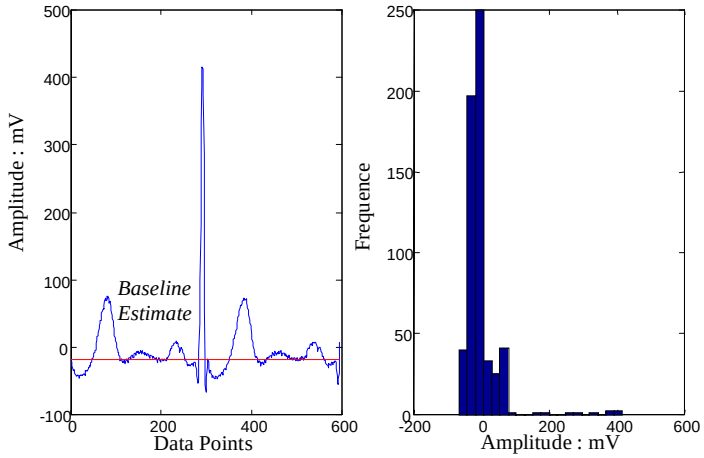

Figure 5-11: *Ambiguous highest ferquency*

This corresponds to a bin that does not fall within the 40% threshold, but is bordered by the next most frequent bins. In the ECG signal, this corresponds to a period of inactivity in the signal with either some systematic variability (such as a non-detected U wave), or moderate amounts of noise. The baseline estimate is the weighted mean (weighted by frequency) of the highest and second highest frequency.

3. All other possibilities (figure 5-11)

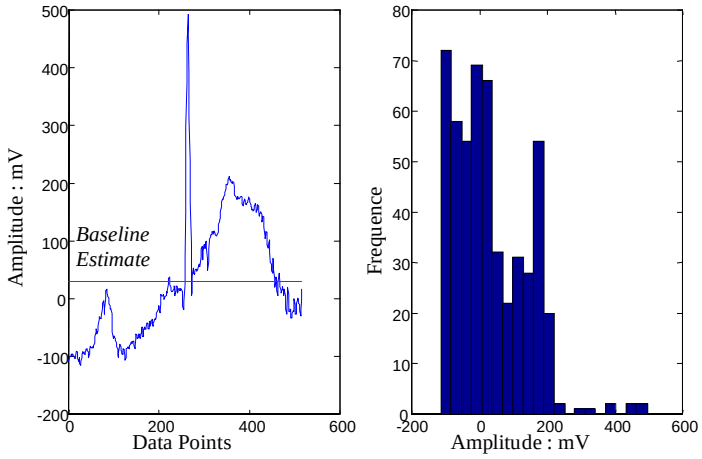

Figure 5-12: *Example of histogram failure: No clear isoelectric regions*

There are two possibilities for non-detection by histogram outcomes 1 and 2. There is no discernible period of inactivity in the window taken, or there is more than one area of minimal gradient that is of a similar gradient. With this in mind, the results of the histogram are ignored and the mean of the sample taken as the best estimate of the baseline.

This last histogram outcome, 3, is a fail safe for spurious signals. The mean however is a notoriously inaccurate measurement of the baseline, owing to the large influence from large R or S waves, raising or lowering the mean respectively. As such, it is preferential to ensure that the window occurs within outcomes 1 and 2.

5. Looping Threshold Increases

This final segment of the algorithm aims to ensure that an estimation of the baseline can be made in all situations. At each step in the loop, the threshold increases by 10% until either the zones pass the accuracy limit, or a maximum threshold of 80% is attained. At this point, if the zones have still not passed the accuracy limit, it is assumed that the pathology has no distinguishable isoelectric line, and the baseline estimate from the histogram will be taken as the baseline. In such a circumstance, the entire window bar the points closest to the R will be taken as inactive. This is necessary to allow noise estimation and baseline interpolation to take place later on.

N.B. Within module 2, there is a final failsafe to ensure that an accurate estimation is passed to back to module 1. If in the second calculation of the active zones (with low threshold), the zones do not pass the percentage criteria outlined before, it is assumed that the waveforms are too large for a correct baseline estimation with the low threshold. In this case the baseline estimate and zones from the previous active zone calculation are used (the first segment marked 2, in module 2 of figure 5-5).

5.6 Interpolation

The final output of the algorithm is a vector containing all of the active and inactive zone information of the entire ECG signal. This active/inactive zone vector (in the form of square

wave with values 0 and 1) can then be applied to the original, entirely unfiltered, ECG signal and its baseline calculated. Various methods have been tested, with the aim of using the information in the inactive zones to correctly determine the variations from baseline wandering, as well as provide a correct isoelectric line. The aim is to do this without deforming the waveforms, or indeed creating new ones. The general idea is to create a baseline signal which can then be subtracted from the ECG signal, leaving a signal with no low frequency noise, and an isoelectric line based at 0.

5.6.1 Flat Line Interpolation

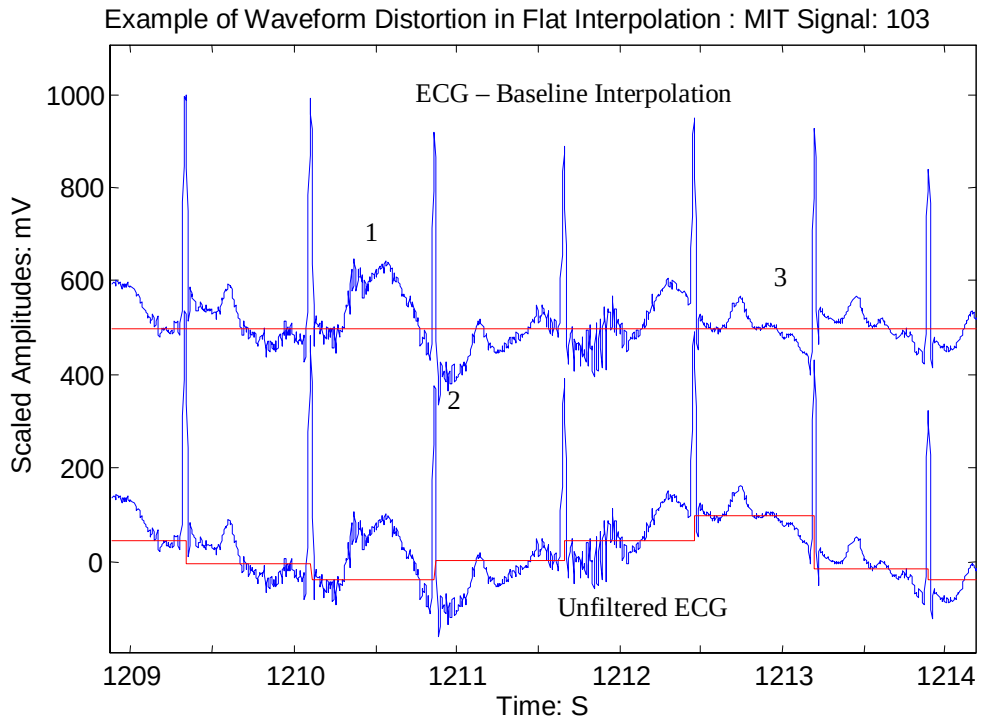

Figure 5-13: *Example of poor adjustment to baseline wandering with flat line interpolation*

This is the most basic method that was used. Again working in the window between two R waves, the baseline is calculated by finding the median of all the points in the inactive zones. This value is then given to all of the points in the window, and is hence a flat line (see figure 5-13). This reduces waveform deformation to some degree by working between R waves, as any jumps in baseline will be accommodated at the point of the R wave, the amplitude of which is of little diagnostic importance. However, it assumes low amounts of low frequency noise, and as such will create waveforms. This deformation is usually greatest around the QRS complex, as this is always active, and so never included in the baseline calculation.

As is seen in figure 5-13, the underlying low frequency noise has not been removed in this interpolation, and a relatively normal signal has been distorted predominantly around the QRS complexes. These can be seen in features 1,2 and 3 on the figure. Each of these could be construed by the HMM as a change in waveform, rather than a normal ECG deformed by low frequency noise.

A new method was therefore proposed to reduce this distortion around the QRS complex.

5.6.2 Linear Interpolation

This method seeks to reduce the sudden changeover between baseline from window to window, as demonstrated previously. As before, the baseline is calculated from the median of the inactive zone points. However, this is then assigned to only those points within the inactive zones, and any points inactive zones between them. The remaining points, i.e. the active zones around the QRS complex, are calculated via a linear interpolation of the incomplete signal. In effect this draws a line between the end of the inactive zones in the precedent window, and the start of the inactive zones in the current window. This is demonstrated in figure 5-14.

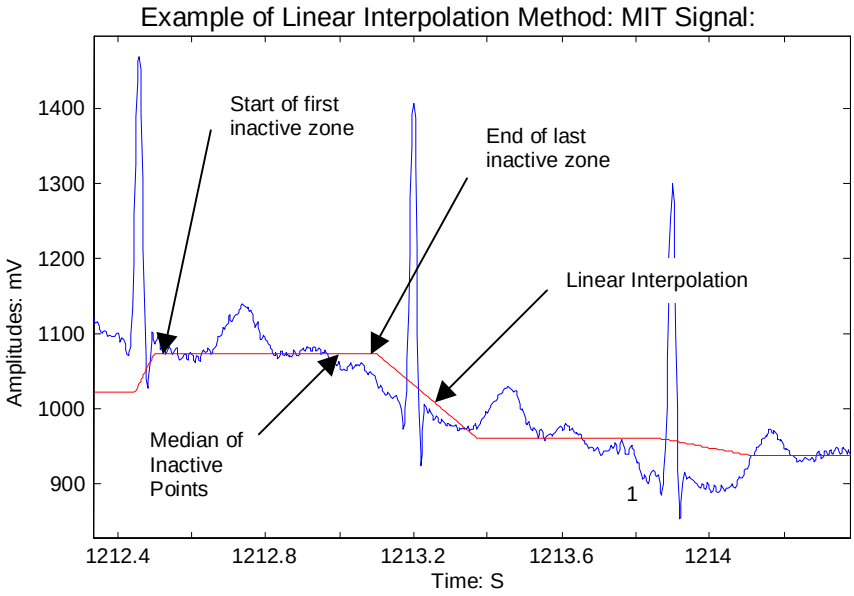

Figure 5-14: *Demonstration of linear interpolation method*

As can be seen in figure 5-14, there is still some deformation around the QRS complex, as the lowest point in the low frequency noise occurs at the end of the R wave (see feature 1 in figure 5-14), and is hence ignored as it lies outside the inactive zone.

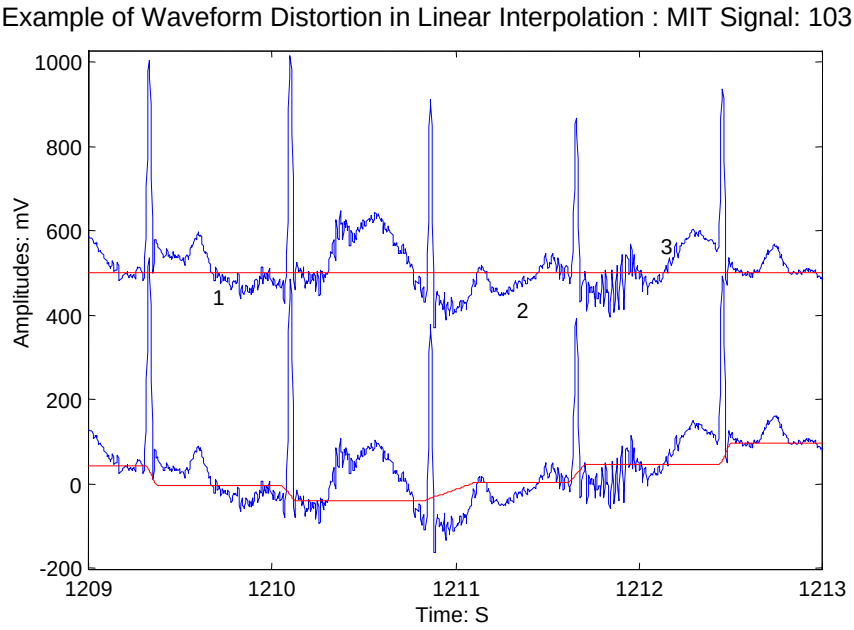

Figure 5-15: *Demonstration of errors remaining from linear interpolation method*

As can be seen again in figure 5-15, though the situation is much improved, there is still deformation of the signal as no account is given to the changes in gradient of the baseline between the R waves (see labelled features 1, 2, and 3).

The final method seeks to take into account all of the information that can be gained from the inactive zones, regarding the behaviour of the baseline wander.

5.6.3 Gradient Interpolation

In this method, the inactive zones are used to analyse whether there is any movement due to low frequency noise. This is achieved by calculating the linear least squares regression of all the points within the window. This line is then given to all of the points in the inactive zones. Between the zones, a linear interpolation is made, as with the previous method.

So as to remove the chance of outliers, caused by the zone detection algorithm, the fit was only made with 80% of the data from the inactive zone. This 80% was on the basis of how close each point was to the median. As such, points on a P or R wave will not be used, while variations in low frequency noise will still be utilised. This allowed a much more accurate fit to be made.

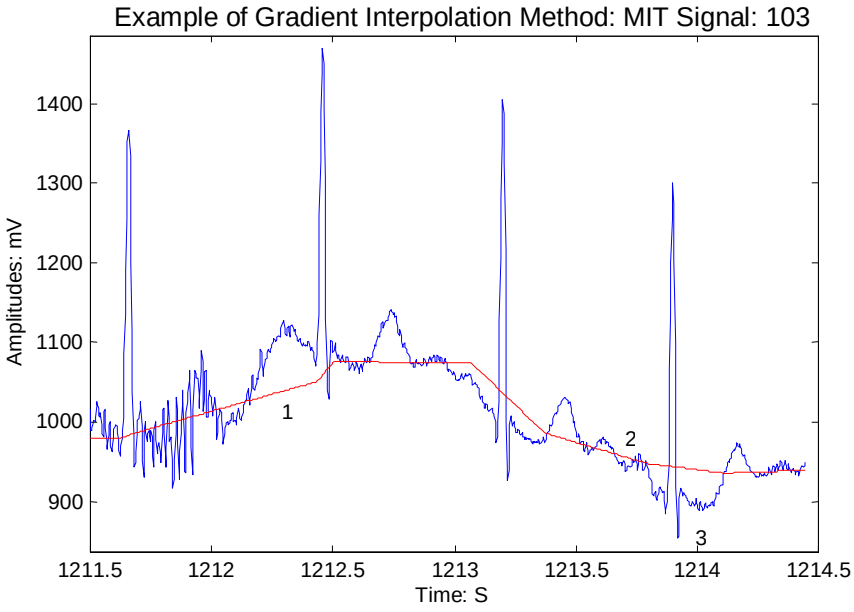

Figure 5-16: *Example of improvements made by using the gradient method*

Feature 1 and 2 in figure 5-16 show that the method has accommodated well to the variations from low frequency noise. However feature 3, demonstrates the continuing problem with large low frequency amplitude changes occurring at the QRS complex, with the linear interpolation passing by these changes in baseline. However as can be seen in figure 5-17, the results have been much improved in the progression from method to method.

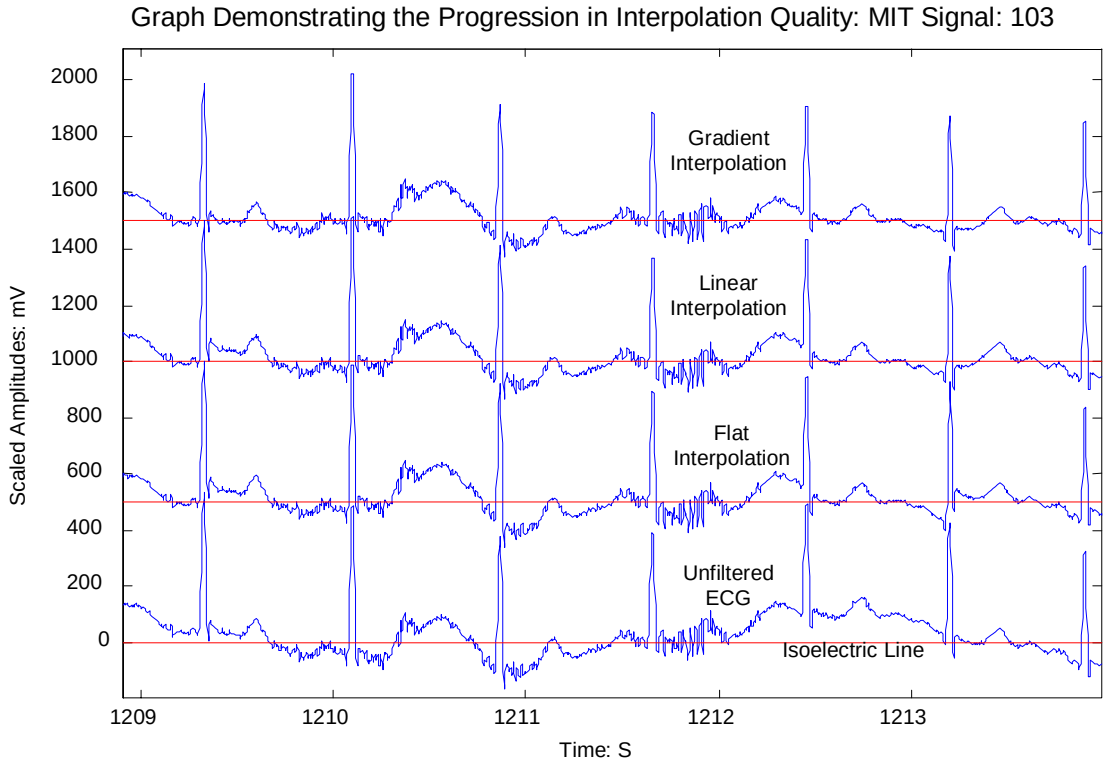

Figure 5-17: *Comparison of different interpolation methods*

This graph represents an ECG signal, with the 3 different types of baseline taken away. Although there are errors with this interpolation, it was decided that the method was adequate

to provide good results for the modelling and HMM phases of the group's work. It is this latter interpolation that will be used to demonstrate the total system with real examples from the ECG database.

5.7 Results

Two types of results are presented in this section:

1. **The accuracy of the algorithm in detecting the isoelectric line.**

Quantitative results were gained in this section using the ECG simulator. These test signals contained no baseline wandering, so this was a direct test of the algorithms accuracy in determining the isoelectric line.

2. **The accuracy of the algorithm in dealing with baseline wandering and isoelectric line detection**

Qualitative results are also presented, using various signals from the MIT database. The sections presented represent areas of the database that contain interesting pathologies of waveforms. These were selected using the cardiologists' annotation, present in each MIT signal.

Results from the ELA test signal are also presented, which provide qualitative results on both of the types of accuracy being investigated.

5.7.1 Results from Simulator

Three variables were altered in this accuracy study: amount of noise added, beats per minute and pathology type. The simulator was set up to produce signals of 100 beats in length, each with a baseline of zero. The baseline algorithm was then used to return a vector **B** of length 100, containing the calculated baseline for each of the 100 R wave-R wave windows. As the correct baseline is zero, the mean of R represents the average error in the baseline calculation. These values have been plotted against each changing variable. In each graph, the distribution of these average errors has been represented by the standard deviation.

Increasing Noise

In this first sample, 50 normal ECG signals were generated each with a frequency of 120 bpm. White noise was then added incrementally with amplitudes between 0 and 20% of the amplitude of the R wave. The baseline of each signal was then calculated, and the average error plotted against noise, with error bars representing the standard deviation. (NB some error bars have been omitted for clarity).

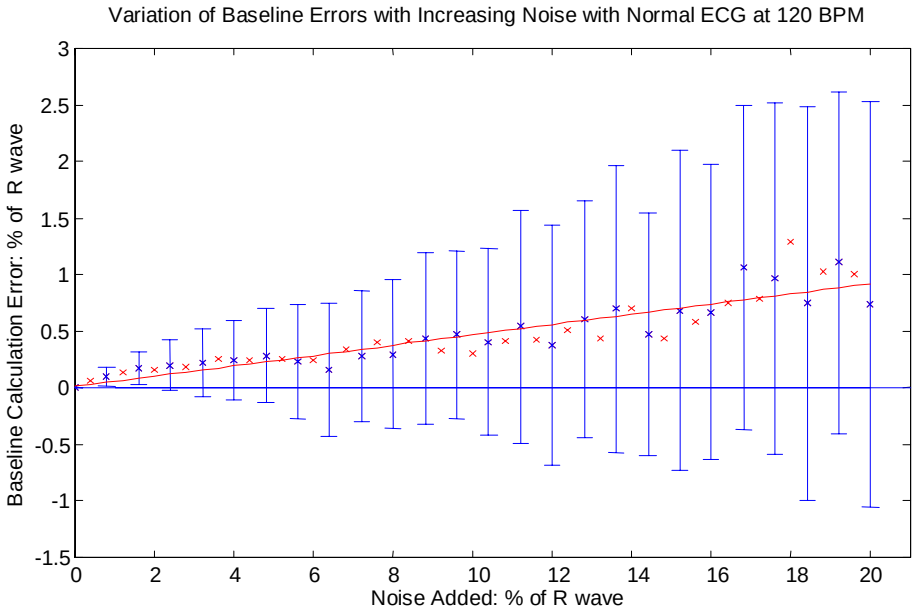

Figure 5-18: *Results from simulator: Accuracy of algorithm with increasing noise*

This graph demonstrates an expected reduction in accuracy as the noise added increases. The maximum error however is around 3%, which is extremely good, considering the large amount of distortion that is present at 20% of noise (see appendix for an example of this signal). From this we are able to conclude that the algorithm has acted independently of the noise added, though there is a slight bias towards a baseline higher up. All of the errors are positive as there are is a small amount of the signal below the actual baseline, thus if any waveforms are included in the calculation, they will have the effect of raising the baseline.

Increasing BPM

The second sample involves 50 normal ECG signals that were generated with frequencies ranging from 80 to 180bpm. A noise of 2% was then added to reduce the artificial nature of the simulator. (Again some error bars have been emitted for clarity)

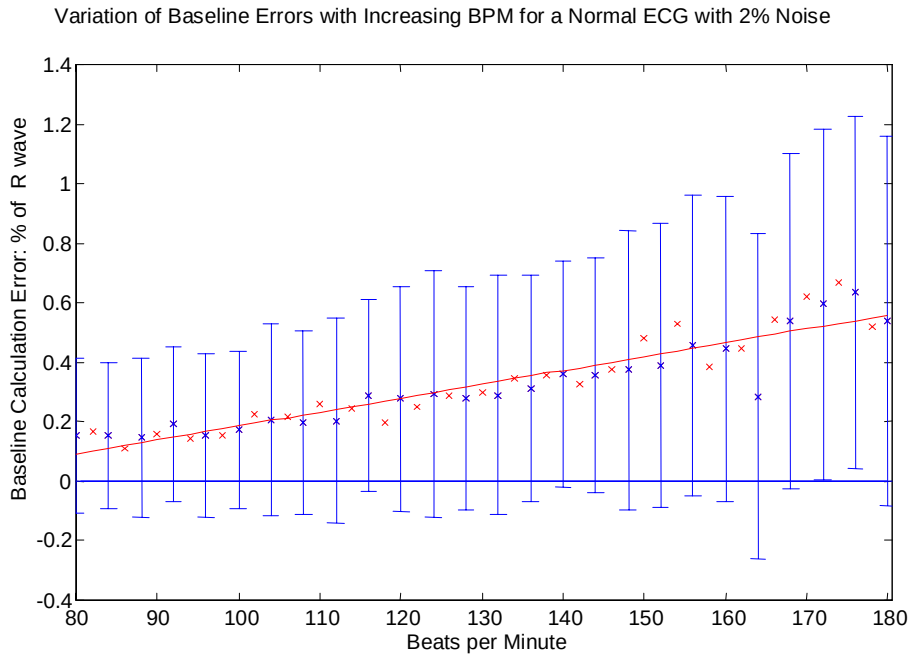

Figure 5-19: *Results from simulator: accuracy variation with increasing heartbeat frequency*

There is some variation in accuracy with increased bpm, though this increase is very small with a maximum error around 1%. As the bpm increases, the amount of signal that is isoelectric becomes smaller, though the results demonstrate that this reduction does not impinge greatly on the accuracy.

Varying Signal Types

This final sample includes 12 signals, each generated from different pathologies (see appendix 3.2 for the key to different signal types). The extremes of each of the two previous variables were taken, to present the worst case scenario to the algorithm (bpm: 180 and noise: 20%).

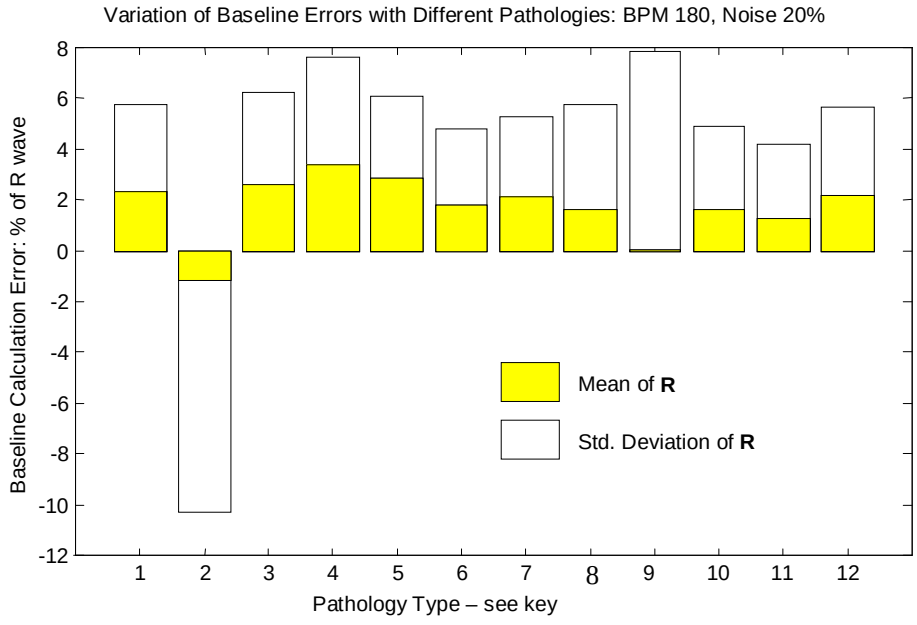

Figure 5-20: *Results from simulator: Variation of accuracy for different signal types*

As can be seen in the graph above, the results for all of the pathology types tested are extremely good, even at the extremes of the variables. The maximum average error observed, is 3%, which is considered extremely small. The largest range of errors is seen in signal type 2, atrial flutter, due to the very limited isoelectric activity in the signal (see figure 7.2 in the appendix).

Comparison with PR baseline detection method

In addition to these qualitative results, the isoelectric line detection accuracy was also measured by comparing the baseline detected using PR segment proposed in section 5.4, and the algorithm described above. Each demonstrate large errors in baseline detection using the PR segment method, whilst in each case, the proposed algorithm is able to correctly interpret the situation, and give a correct baseline estimation. These results can be found in appendix A2.2. These were tested without baseline wandering (prefiltered using high pass filter), and thus represent only inaccuracies with isoelectric line detection.

Summary of simulated results

The accuracy levels are very high considering the wide variety of signals tested. All of the errors were below 5% even at the extreme conditions, well below the amplitude of the important features (The P wave has an estimated amplitude of 10-15% of the R wave). The dispersion of the results (up to 10% in type 2) can be explained in part by the limited number of points over which the noise was added, hence risking a noise with a non-zero mean. At the 20% level, this becomes highly possible and will therefore introduce a bias. Indeed, at the 20% level, almost all of the features in the signal are lost, and the algorithm is measuring noise.

5.7.2 MIT-Database Results

Throughout the project, the algorithms were tested against the MIT arrhythmia databases. There is no information within this base concerning a correct isoelectric line, so a quantitative calculation of the accuracy is not possible. The database does however give the location of interesting pathologies and regions of noise, which have been used here to give an indication as to the quality of the algorithm and possible improvements.

In each graph, the original unfiltered signal is shown with the interpolated baseline (upper signal), as well as the unfiltered signal with this baseline subtracted. To show the two traces on the graph, this second signal has been shifted upwards slightly though the relative scales remain the same.

Correct Analysis

With relation to this sections aims, three ideas can be shown here: the ability of the algorithm to work through both high and low frequency noise, and the applicability to all types of pathology present. These results have been chosen to be representative of the aims of the project, and have been placed in appendix A2.1. Results include the baselines of various difficult pathologies, as well as both high and low frequency noise variations.

Limitations

It is very encouraging to note the high level of accuracy achieved by the algorithm in most cases, a fundamental objective of the project. However a small number of cases of incorrect analysis are described below.

Lowering of QRS complex

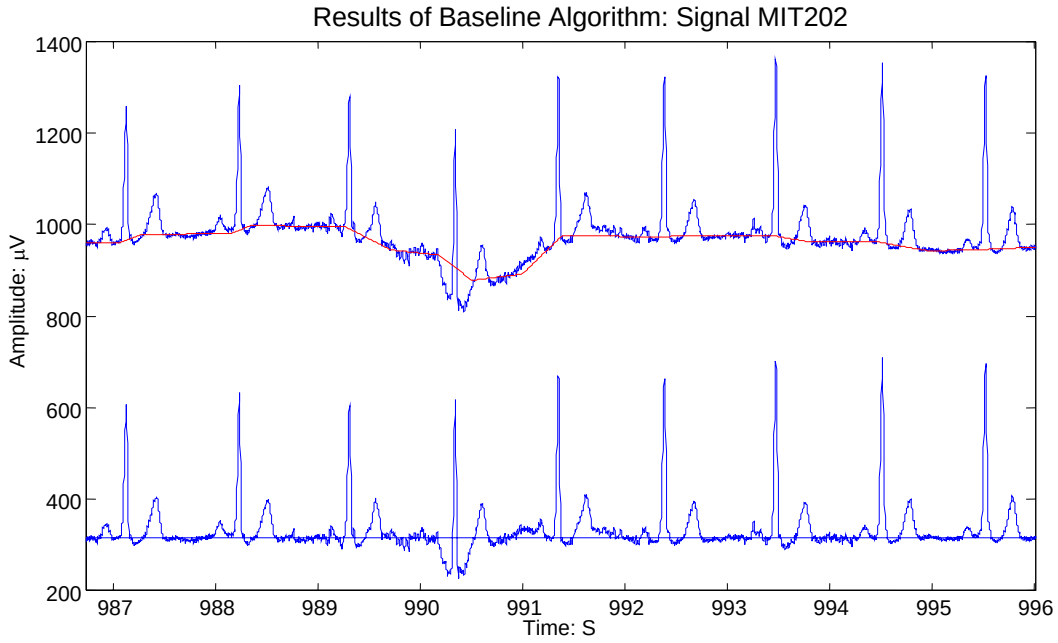

Figure 5-21: *Results from MIT database: Lowering of QRS complex due to linear interpolation across active zone*

The most significant anomaly in the signals is the lowering of the QRS complex due to low frequency noise. As shown, this results from the linear interpretation across the QRS complex, correctly joining up the two detected regions of inactivity. However, in the algorithm no account has been made of information within the active zones, hence any dramatic changes in such a region will not be detected. This has been resolved to a certain degree at the modelling stage, with the addition of several Gaussians of large width to the wavelet library.

Ventricular Fibrillation

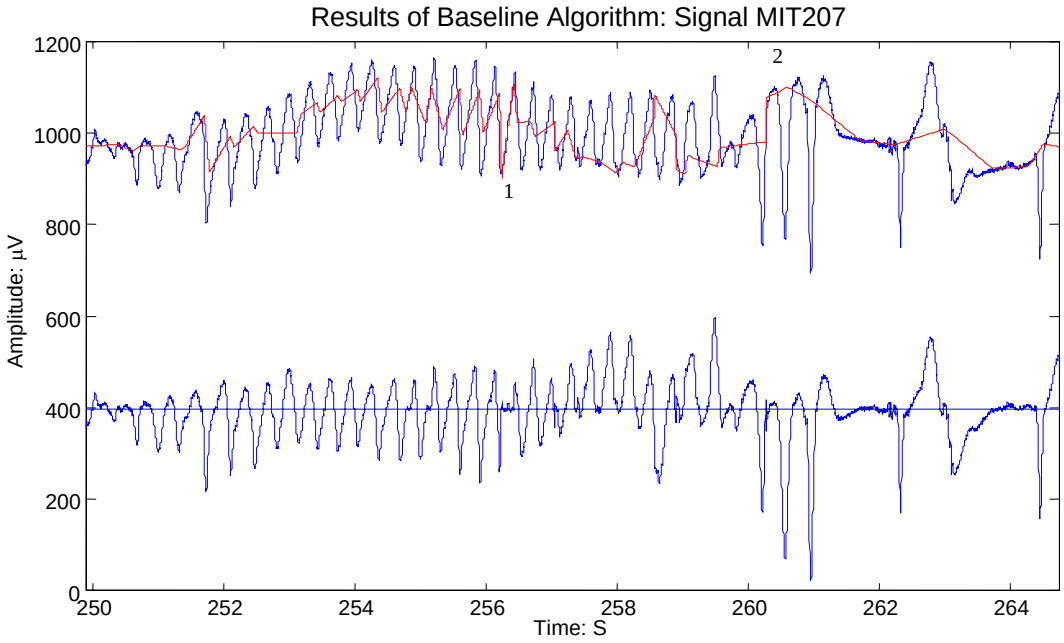

Figure 5-22: *Results from MIT database: Signal distortion in ventricular fibrillation due to poor interpolation*

Two errors are visible in this graph in features 1 and 2: The first is due to the interpolation process rather than the algorithm. Here the pathology has been correctly interpreted as one without an isoelectric line (i.e. fibrillation), and the entire R-R region has been made inactive. These results are then interpolated via the linear regression, giving the line up the slope of the signal. As can be seen however, the waveform deformation is less of a problem, as it is meaningless to calculate relative amplitudes or waveform transformations in this region; it simply needs to be recognised.

Feature 2 involves an error in the histogram detection, with the small flat region at the crest of the signal being interpreted as isoelectric, and thereby as baseline.

Deformation of ST region

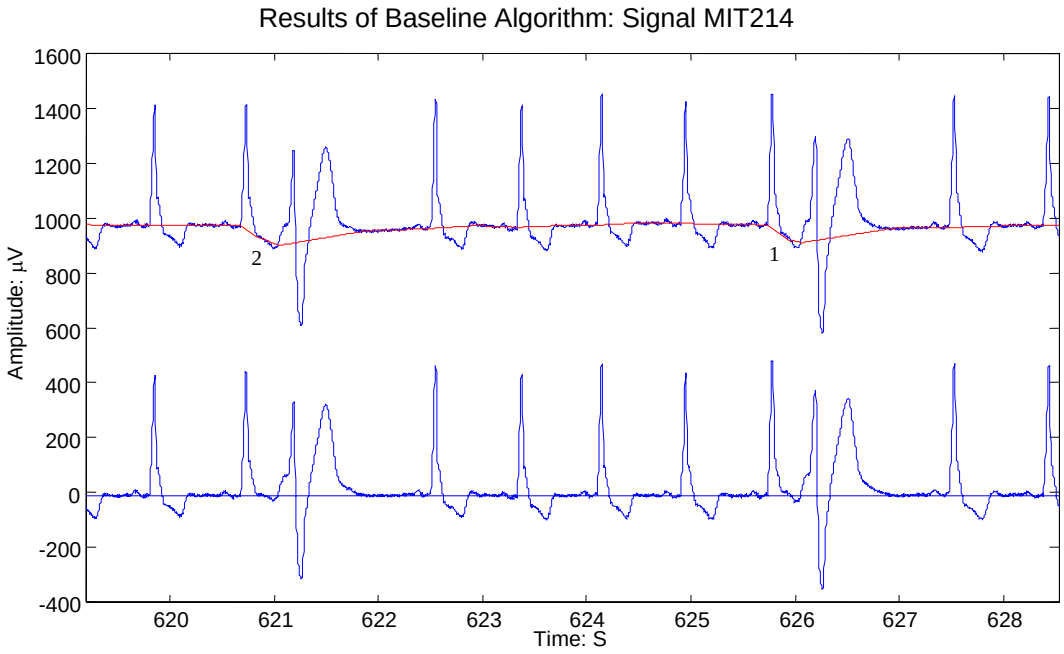

Figure 5-23: *Results from MIT Database: Deformation of ST region in ventricular extrasystole due to interpolation method*

As can be seen in the features 1 and 2, the lowered ST segment is not represented in the treated signal, owing again to the calculation of the gradient in this region, removing this effect. This is a minor effect again caused by false interpolation, rather than the algorithm

5.7.3 ELA Test Signal

The ELA test signal was used to present the worst case scenario, with a difficult to analyse pathologies and both high and low frequency noise. For the majority of the signal, the baseline is correctly found, with extremely good precision.

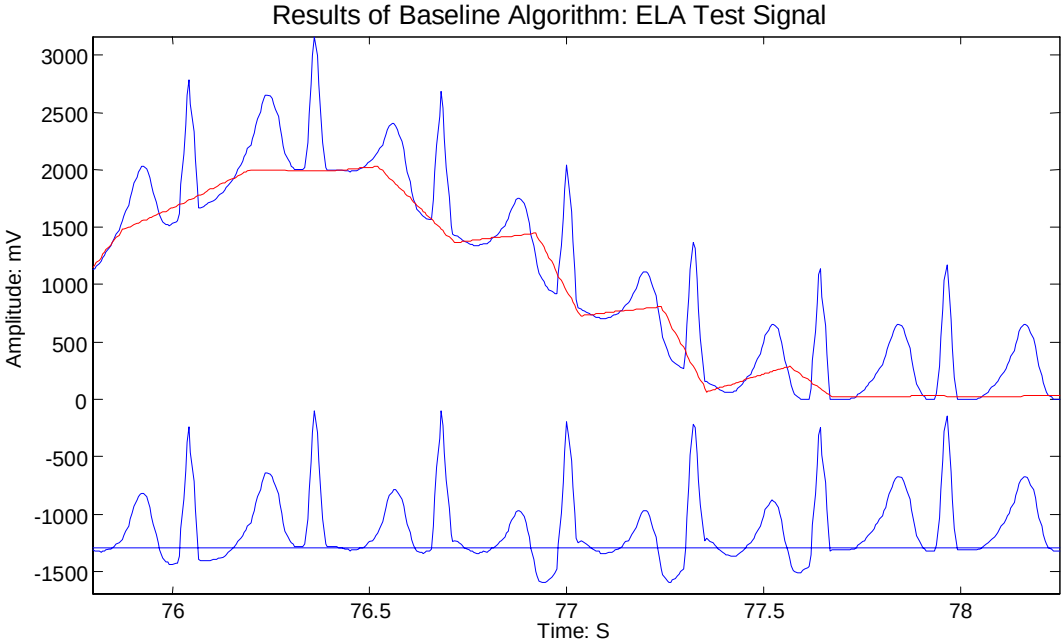

Figure 5-24: *Results from ELA test signal: poor baseline wandering approximation due to linear interpolation method*

However, in the regions of the added low frequency noise, an incorrect interpolation is seen, with some waveform deformation. This is due to the signal being interpreted as one without isoelectric line (caused by the high amplitude waveforms) and so the entire R-R region is made inactive. The resulting interpolation is thus false and not applicable.

High Frequency Noise

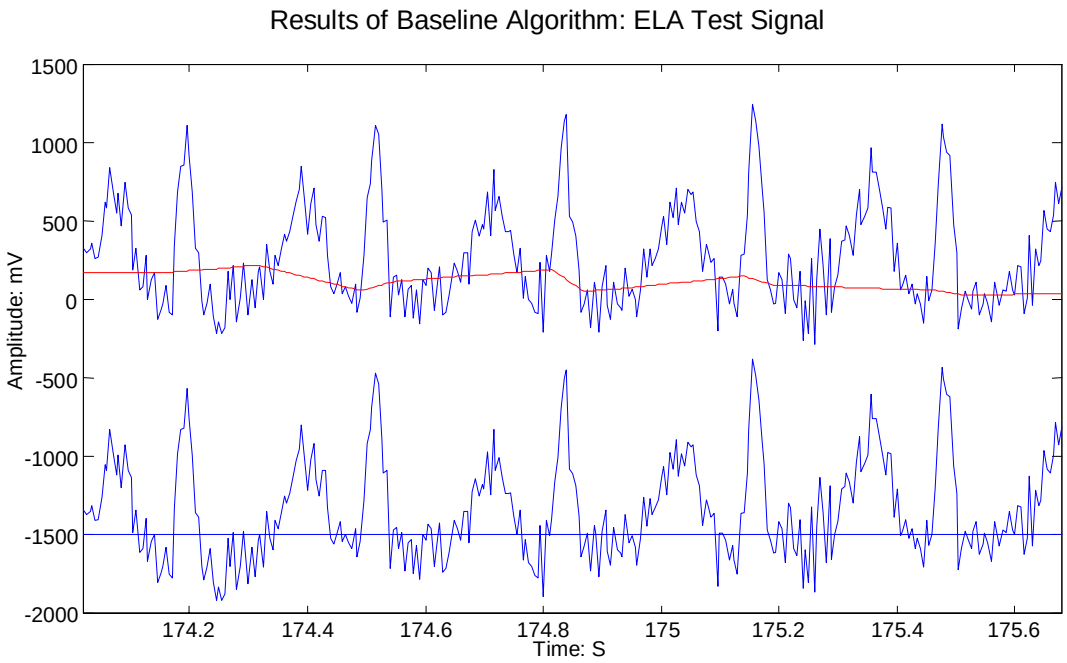

Figure 5-25: *Results from ELA test signal: reduced accuracy of detection in region of high frequency noise*

Within the region of added high frequency noise, there is a reduction in baseline quality towards then end of the signal where the noise is greatest. This is expected as the noise is of high enough amplitude in this zone to be considered as an entirely new zone and to be able to pass the threshold. Thus there are no recognisable isoelectric segments and the algorithm has made the entire signal inactive. Again it is the interpolation that is at fault, as a flat line through the median of this, would give a correct baseline.

5.7.4 Summary of Difficulties from MIT Results

- **Incorrect Interpolation**

As is seen in the results, the use of the line of the best fit as the baseline has its limitations, especially evident in the case of ventricular fibrillation (see figure 5-22). In all these examples, the inactive zones have been correctly picked, but the line of best fit through this zone is not applicable, leading to a deformation of the signal.

- **Low Frequency Noise at QRS**

Again in the cases proposed above, the inactive zones have been correctly defined, however if the trough of the low frequency noise falls on the QRS complex, it becomes difficult for the baseline to be interpreted, and results in false waveforms.

5.8 Summary

Method

- Zones selected by variable thresholding of scaled importance values
- Various failsafes are utilised to ensure accurate zone assignment (active/inactive)
- Information within inactive zones then interpolated to give an estimate of isoelectric line

Results

Isoelectric Line Detection Results

- Simulator – High levels of accuracies for all signal types, heartbeat frequencies and noise levels
- PR Segment Comparison – The algorithm is shown to give far more accurate results than the PR method, for all variation types
- ELA Test signal- For regions without baseline wander, the algorithm is seen to produce very high accurate results, despite the difficult pathology presented

Baseline Wandering Results

- The majority of results from the MIT database are shown to interpret the large variations in baseline to a high degree of accuracy
- Cases are shown where the interpolation method is false (e.g. lowered QRS)
- Poor results shown in tests on ELA test signal in regions of sharp low baseline wandering

5.9 Conclusions and Improvements

The first results from the simulator demonstrate the very high accuracy of the algorithm described in this chapter, when used on a signal with limited low frequency variation. The errors found are negligible in comparison to the amplitudes of the important features being analysed, and thus there is very small risk of waveform deformation.

The results from the MIT database, demonstrate that although extremely accurate for almost all pathologies, there are problems regarding applying the detected inactive zones to the unfiltered signals. These problems are associated specifically with how to interpret the baseline when it has a gradient due to low frequency noise. Whilst the line of best-fit works for the majority of cases, it does cause difficulties when signals with sharp variations at the QRS complex are considered.

The author has assumed here that the reasons for inaccuracies are not due to zone detection. This is primarily true, as the increased number of zones due to inflections without maximums does not greatly affect the distinction of a correct baseline.

Two possible improvements are proposed.

The first most direct approach would be to improve the interpolation of the active zones. This would involve analysing information from the active zones, regarding their underlying

gradient, and not just performing the straight or linear interpolation across them. This could be investigated by reference to the positions of the inflection points that define each of the zones. A further direct improvement would be to not use the gradient interpolation in windows where it is decided that there are no inactive zones, such as in the case of fibrillation.

A second possible improvement would be to distinguish and separate the two processes: removal of baseline wandering (thus low frequency noise), and calculation of the correct isoelectric line. The algorithm proposed here is not optimal to do both, demonstrated by the extremely high level of accuracy found in the simulator results, and various difficulties found in the MIT databases. This will require a separate method to remove the low frequency component of the spectrum, and then the baseline algorithm to find the inactive zones. This will result in no need for the interpolation and hence the correct straight-line path through these isoelectric segments, can simply be given by the median. This will also mean a decreased calculation time, as there will be fewer occurrences of the signal being passed to the second and third modules in the algorithm.

The accuracy with which the baseline wander is removed would not have to be great, as any large variations in baseline would be treated as erroneous low frequency noise, and thus not analysed. As such a high pass filter with a reduced cut off frequency would probably suffice, thus eliminating any risk of waveform deformation.

In conclusion the project aim of calculating the isoelectric line has been achieved to a very high level of accuracy, as shown in the results from the simulator. However, the system has proved to not be optimal for the projects second aim, that of baseline wandering removal, though the regions of inactivity in this signal are always robustly defined.

Chapter 6. High Frequency Noise Estimation

Both high and low frequency noise levels need to be estimated in the system, to ensure that the models applied in the next step of the group's work are accurate. The estimation must be done on a window by window basis, as it is within this unit that the next stage of wavelet analysis will be carried out. It is worth noting the distinct difference here between the group's aims and those of other workers. Whilst many researchers aim to remove noise via numerous involved filtering processes, this group wishes to retain the signal as it is, and simply not analyse the areas where the noise is deemed too great. This reduces the risk of misdetection from filtering distortion, an important point given the frequency spectrum of the ECG.

6.1 Motivation

In the next step of the group's project, each heartbeat is modelled using Gaussian wavelets (see section 3). The combination of these wavelets allows any given signal to be accurately modelled, though there is no account given to the quality of the model. Hence some of the chosen wavelets may represent large amplitude noise and thereby give a false wavelet into the HMM system. The HMM works only from the parameters of the wavelet, thus allowing this noise to be interpreted as a waveform. This is a rare occurrence however, due to the robustness of the wavelet system in place, and the ability of the HMM to pick out errors.

Perhaps more important is the number of calculations required in a long ECG signal. It is a time consuming process to model each waveform and then optimise each with a neural network. In a 30 minute ECG signal however, there may only be 5 or 6 different types of heart sequence (PQRST), with very limited changes in the optimised wavelets needed for accurate modelling. This is because, whilst a patient may have numerous different waveform sequences, these will generally occur with the same characteristics. With this in mind, it is possible to build up libraries of the patient's different waveform sequences (e.g. ventricular extra systole and normal SA regulated heartbeat), against which each new signal is checked.

If the signal in the current R-R window does not correspond to any in the libraries, it undergoes wavelet modelling and neural network optimisation and a new library is created. However, if there are sequences with high levels of noise, the correlation between it and that of the library sequence will be reduced. This leads to excessive numbers of libraries being created, and a vast increase in the time needed for calculation. The aim is to remove these high noise heartbeats from the ECG signal and not model them.

6.2 Noise Hypothesis

Again, the ECG signal is split into windows, with the start and points being defined by successive R wave. This are labelled by i where i = 1 to N-1 and N is the total number of heartbeats (Hence window (5) is measured from R-wave (5) to R-wave (6)). Within these

windows, we assume that the noisy ECG signal, $\mathbf{Y}^i$, can be separated into three separate entities, the ECG signal $\mathbf{X}^i$, low frequency noise $\mathbf{L}^i$ (due to baseline wandering) and high frequency noise $\mathbf{H}^i$ (due to muscular contraction), where i = 1 to N-1 and N is the total number of detected heartbeats in the entire signal.

$$\boldsymbol{Y}^i = \boldsymbol{X}^i + \boldsymbol{L}^i + \boldsymbol{H}^i \qquad \textit{Equation 6-1}$$

As the measured ECG signal is discrete, each of these entities have n^i points, where n^i is the number of points within window i. We make the hypothesis that the high frequency noise within each window is Gaussian and so has a normal distribution with a mean of zero, and a variance of σ^2.

$$\boldsymbol{H}^i \sim N(0, \sigma^2) \qquad \textit{Equation 6-2}$$

By making this hypothesis, we are able to use an estimate of σ, to characterise the amplitude of noise within each window. As the distribution is normal, the larger the overall amplitude of noise, the larger the value of σ will be. Over various tests later on in the project, the noise was indeed found to be Gaussian in nature, especially considering the large sample of points that were being considered in each sample.

6.3 Testing Methodology

As the MIT database does not provide substantial information on the amounts of high frequency noise, the simulator has again been used to provide quantitative accuracy estimates. In each graph involving the simulator, 20 different test signals were created, and as with the baseline detection results, one of three variables was altered for each signal: σ of noise added, heartbeat frequency or signal type. The noise added was created using the Matlab command randn, giving the distribution N(0,1), and the σ was adjusted by multiplying the noise by a percentage of the R wave. Hence for an R wave with height 1, a 20% noise is distributed as N(0,0.2).

The amount of noise added was then compared with that estimated by the different methods. With each test signal, all of the heartbeats had the same three characteristics as outlined above, and an estimate of σ in each window (i) was returned. The mean noise amplitude was then calculated, along with the standard deviation, over all the windows, giving a better statistical overview of the systems accuracy. These mean values are presented on all of the graphs against the changing variable, with the standard deviation as the errorbar to give an idea of the range of results.

6.4 Methods

Each of the three methods aim to extract information on H^i in equation 6-1, and each method varies in the way it removes the effects of the other 2 entities.

6.4.1 Standard deviation of inactive zones

This initial method first hypothesises that the effects of L^i are negligible over the window, and that X^i is very small in the inactive regions calculated in the previous chapter. This will mean that variations in amplitudes of points within the inactive zones are due solely to high frequency noise, thereby allowing a direct estimate of σ to be made. This assumes that the inactive regions are ideally isoelectric.

Initially, to estimate the value of σ, the unbiased standard deviation, s, was used in each window, i.

$$s_i^{\,2} = \frac{\sum_k \left(Y_k - \overline{Y}_i\right)^2}{n-1} \qquad \textit{Equation 6-3}$$

where k = Inactive points within window(i), n is the number of inactive points and $\overline{Y}_I$ the mean of the points within the inactive zones. However, the dramatic effect of outliers was noted and an adjusted standard deviation was used, using the median in place of the mean of the signal.

$$s_i^{\,2} = \frac{\sum_k \left(Y_k - median(Y_i(I)\right)^2}{n-1} \qquad \textit{Equation 6-4}$$

Where Y_i (I) are all the inactive points within window(i). This reduces the bias towards larger values of the signal that are included in the inactive zone by error. (see figure 4-17)

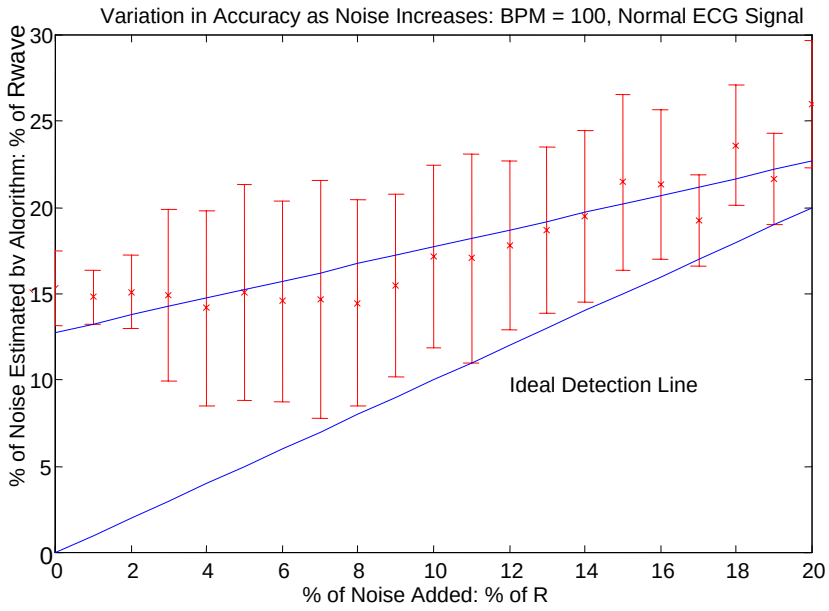

Figure 6-1: *Variation in estimation accuracy with increasing noise, using first method*

Whilst the results from the simulator for a normal signal were shown to increase, (figure 6-1) on application to a real situation signal, (i.e. MIT database), the hypothesis of an ideal isoelectric line was found to be unsupportable. Reasons for this include:

- Poor zone detection causing higher amplitude parts of waveforms to be included with noise
- Waveforms not detected e.g. U wave causing large outliers
- No discernible inactive zone – In this case the baseline algorithm will select a zone sufficient to estimate the baseline and hence include parts or the entirety of waveforms.
- Underlying low frequency noise causing variations in isoelectric line

Each of the above reasons will give a large standard deviation for even the 'cleanest' signals. Mathematically this is unacceptable. These problems were accentuated with the Marr-Hildreth algorithm, as by choosing the points of inflexion as zone cut off points, it is ensured that at least a certain part of the waveform will be within the inactive zone.

6.4.2 Model - Low Pass Filter

A better method would be to first approximate the signal and subtract this from the full signal, giving only the noise present. Hence, we wish a signal that will give an estimate of both $\mathbf{X}_i$ and $\mathbf{L}_i$ in equation 6-1. This will then be subtracted from $\mathbf{Y}_i$, to leave only $\mathbf{H}_i$. This allows $\mathbf{H}^i$, to be estimated in the same manner as previously.

Initially the model used was a low pass digital filter (Butterworth) of the noisy signal, with a cut off above the maximum frequency in the QRS complex (around 100Hz). This produced much better results than the previous method, especially when applied to real signals.

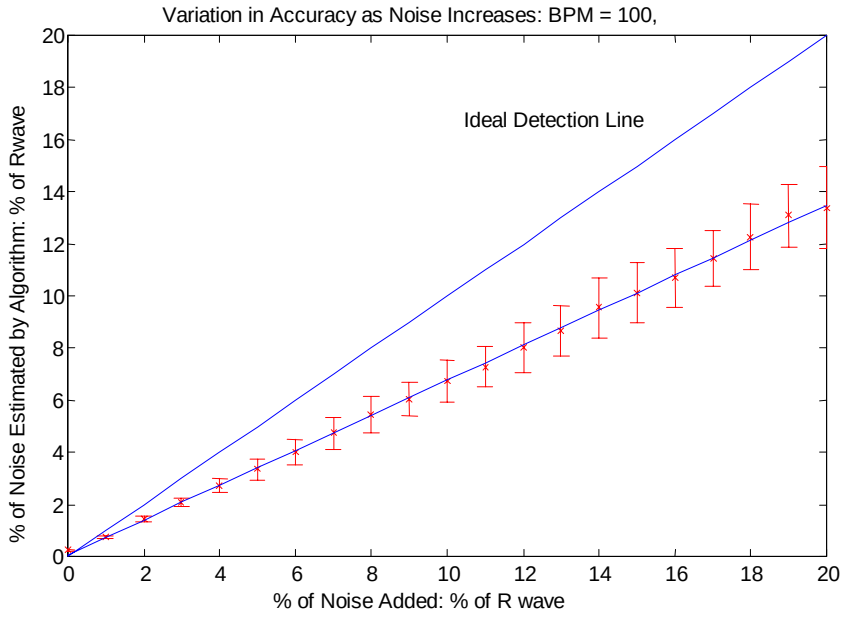

Figure 6-2: *Noise estimation for varying noise, using low pass filter to estimate HF noise and signal*

A principle feature of this graph is the line of best fit passing through zero, hence the error is zero in the optimum circumstance, a great improvement on the previous method. However, the accuracy of the estimate of σ, was found to reduce linearly, as a function of the noise

added. This is seen in figure 6-2, with the estimate of σ drifting further away from the ideal estimate of σ. This is explained by the model (the filtered signal) approximating the noise in addition to the signal. As such, the difference between the model and signal is reduced in areas of noise. It is not possible to decrease further the cut- off frequency, without risking a complete deformation of the QRS complex.

3. Model – Gaussian convolution

This problem of under estimation was reduced by applying the Gaussian filter used in the zone detection algorithm (see figure 6-3). This provides a much better model of X_i and L_i, with a greatly reduced amount of noise modelling. However, the filter does deform the signal at high frequencies, such as in the QRS complex. Hence if the inactivity zone falls across part of the QRS, it is possible to get large outliers. A failsafe was introduced to remove this problem, with all points 40ms before and after the R wave being placed in an active zone. This effectively removes all of the important high frequency waveforms (QRS) from the signal.

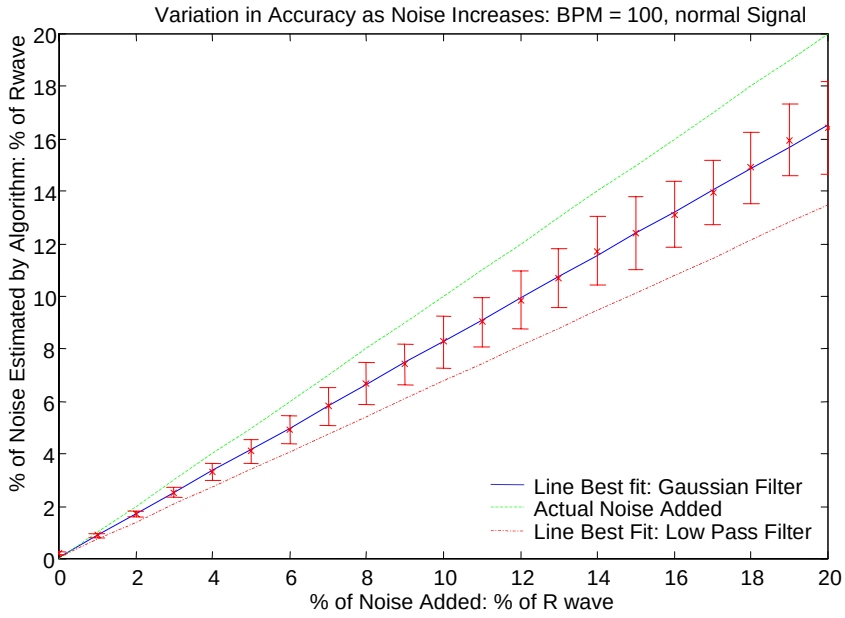

Figure 6-3: *Comparison of noise estimation using Gaussian filter and low pass filter, for varying noise levels*

This however requires a valid standard deviation (σ_g) to be chosen in the Gaussian, convolved with the signal.(see section 4.3) If this is poorly chosen and is too high, the model will rise much slower than the signal in areas of amplitude change, and thus increase the modelling error. Equally if too low, too much of the noise will be modelled and the same problem with the high frequency is found. In the example in figure 6-3, the filter is seen to over estimate the noise, thus producing a lower calculation of the noise. In addition, the results demonstrate a very low dispersion of the estimate, implying that the system used is very robust, in comparison to the large amount of noise added at the extremes.

The results were an improvement on both previous methods and were felt to give an accurate measurement of noise in the simulated signal. When compared with the real signals in the MIT database, the system was able to pick out the windows that should not be measured.
This method was then tested again quantitatively with the simulator, and qualitatively with the MIT database, and the ELA test signal.

6.5 Results from the Simulator

The aim of using the simulator was to produce quantitative results regarding the accuracy of the noise estimates when all 3 of the variables discussed in the previous chapter were altered. It was demonstrated in the method, that the noise estimates were extremely accurate when the amount of noise added was varied. It remains to demonstrate results from altering the heartbeat frequency, and the signal type.

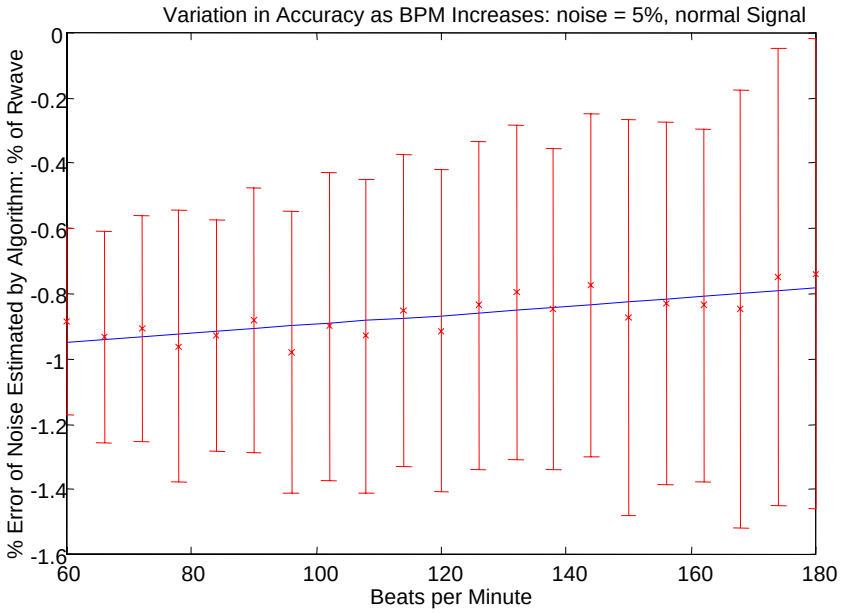

Figure 6-4: *Variation in accuracy of noise estimation with increasing heartbeat frequency*

6.5.1 Heartbeat Frequency

Again 20 signals were created with heartbeat frequencies ranging from 80 to 180bpm. A noise of 5% was then added to each, and the system was used to estimate this noise. Figure 6:4, represents the accuracy to which this value of 5% was estimated (hence, the y axis represents estimated σ - 5%).

The results demonstrate that there is negligible dependence of accuracy with bpm, with the difference in estimates remaining at around 0.8%, as predicted by the curve in figure 6:3. In fact, it is shown that the accuracy increases slightly as the bpm increases. This indicates that the underestimation of the Gaussian filter has been reduced at higher heartbeat frequencies.

6.5.2 Signal Type

In the final experiment with the simulator, 12 different signals were created with both the noise, and bpm constant, and a varying signal type.

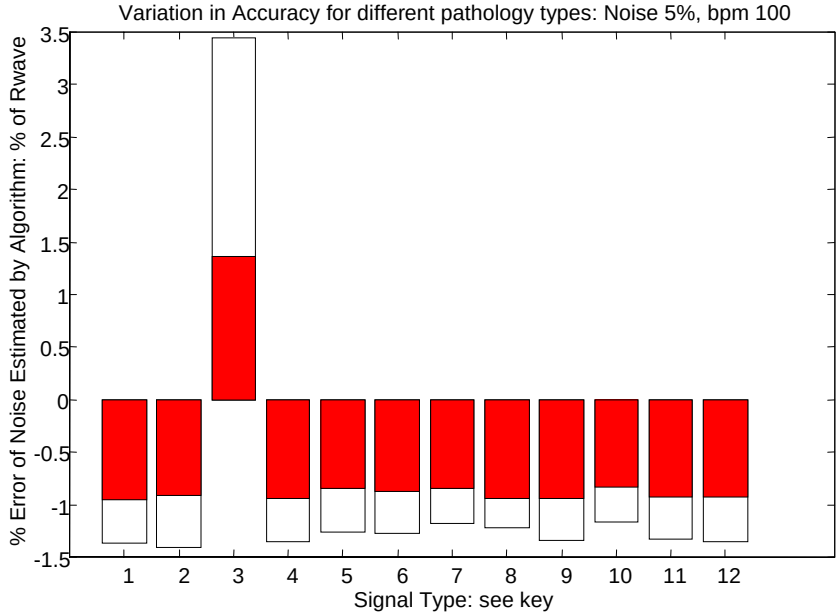

Figure 6-5: *Variation in accuracy for different pathology types, for constant noise and bpm*

As can be seen in figure 6-5, there is very little variation in accuracy, all remaining at the predicted 0.9% level. The only different trace, is again signal type 3, atrial flutter. This has a higher than average error in noise estimate, owing to the difficulty the Gaussian filter has in determining the signal type with such the high frequency P waves as seen in the signal key in appendix A3.2.

6.6 Results from MIT Database

Again there are no quantitative results possible from the MIT database. However, several results are shown in appendix A2.3, with the spectrum of noise demonstrated, and the corresponding signals of various features within that signal.

6.7 Results from ELA Test Signal

The ELA test signal consists of three distinct regions, no noise, medium amplitude noise, and high amplitude noise. In addition, there are areas of extreme low frequency variations and higher heartbeat frequency.

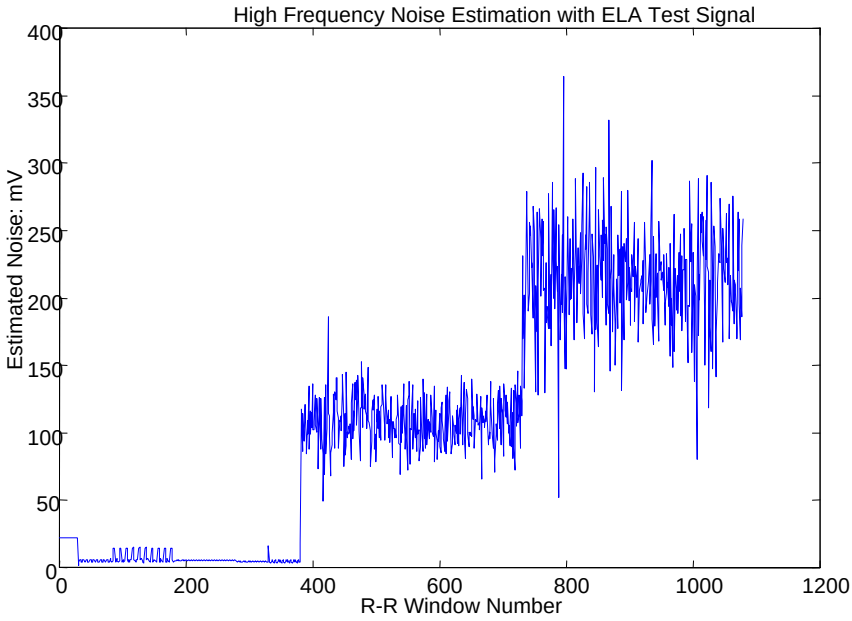

Figure 6-6: *Noise estimation in ELA test signal, demonstrating the different levels of noise added to the signal*

As can be seen in figure 6-6, the system has clearly distinguished the three different levels of noise. Important features of the system are demonstrated in this signal:

- There is no deviation in areas of low frequency noise
- A correct zero estimate given in the areas with zero noise (hence no bias from signal)
- There is no bias from an increasing heartbeat frequency.

(see appendix 3.3 for an indication of the types of noise within the ELA test signal)

6.8 Summary and Conclusion

Key points:

Method

- Noise estimate made on a window by window basis
- Gaussian filter used to remove X_i and L_i from equation 6-1
- Estimate of H_i, made within previously identified inactive zones

Results

- Estimated noise in the simulator very close to true values
- Very low dispersion of results
- Results linear
- No effect of changing signal type or heartbeat frequency

In conclusion, the results demonstrate a high level of accuracy in determining the amount of high frequency noise in a given ECG signal. Possibly the most important point is the linearity of the estimate in relation to the actual amount of noise present. Hence, if there is twice as much noise in the signal in one place than another, the noise estimate will also have this same

factor. This will allow a meaningful threshold to be applied to the results, enabling the problem ECG windows to be removed from the modelling section of the report.

6.9 Areas for Further Study

For a full optimisation of this method, further areas need to by studied. These include:

- Investigation into a suitable threshold level

Now that the noise has been estimated, these values must be thresholded, such that only analysable signals are passed to the modelling phase of the group's work. This must take into account that if too many of the heartbeats are removed from the process, the accuracy of the pathology diagnosis will be reduced, as well as the initial motivation of this study as outlined in section 6.1.

- Investigation into effects of σ choice in Gaussian filter

A full study needs to be made as to the optimal choice of σ, taking into account the different variation in sampling frequencies.

Bibliography

[1]WHO, “Cardiovascular Diseases,” www.who.int/ncd/cvd/ (1999).
[2]J Adamec and R Adamec, *ECG Holter Manuel D'Interprétation Electrocardiographique*, 1 ed. (Médecine & Hygène, Paris, 2000).
[3]MIT-BIH, “Physiobank ECG Signal Databases,” http://www.physionet.org/physiobank/ (2001).
[4]D.A. Coast, R.M. Stern, G.G. Cano *et al*., “An Approach to Cardiac Arrhythmia Analysis Using Hidden Markov Models,” IEEE Transaction on Biomedical Engineering **37** (9), 826-836 (1990).
[5]R Dugad and U B Desaj, “A Tutorial on Hidden Markov Model,” Technical Report Indian Institute of Technologie **SPANN - 96.1** (1996).
[6]ELA, “Web Site,” http://www.elamedical.com (2001).
[7]M.E. DeBakey and A.M. Gotto, *Le Coeur en Action* (Sanofi Synthelabo, USA, 1997).
[8]D Dubin, *Lecture Acceleree de l'ECG*, Maloine ed. (Paris, 1999).
[9]B Vidakovic and P Müller, “Wavelets for Kids,” (1991).
[10]R Lopez, “Advanced Engineering Mathematics,” (Addison-Wesley, 2001), pp. 756-762.
[11]D Marr and E Hildreth, “Theory of edge detection.,” Proc R Soc Lond B Biol Sci **207** (1167), 187-217 (1980).
[12]Babaud-J, Witkin-AP, Baudin-M *et al*., “Uniqueness of the Gaussian kernel for scale-space filtering,” IEEE-Transactions-on-Pattern-Analysis-and-Machine-Intelligence **PAMI-8** (Jan.), p.26-33 (1986).
[13]M.P. Royo, “Baseline Wandering Removal: A Comparitive Study (Spanish),” presented at the XVI Annual Conference of the Spanish Society of Biomedical Engineering, CASEIB'98, Valenica (Spain, 1998 (unpublished).
[14]CR Meyer and HN Keiser, “Electrocardiogram baseline noise estimation and removal using cubic splines and state-space computation techniques.,” Comput Biomed Res **10** (5), 459-70 (1977).
[15]JN Froning, MD Olson, and VF Froelicher, “Problems and limitations of ECG baseline estimation and removal using a cubic spline technique during exercise ECG testing: recommendations for proper implementation.,” J Electrocardiol **21 Suppl**, S149-57 (1988).
[16]V Shusterman, SI Shah, A Beigel *et al*., “Enhancing the precision of ECG baseline correction: selective filtering and removal of residual error.,” Comput Biomed Res **33** (2), 144-60 (2000).
[17]L Sornmo, “Time-varying digital filtering of ECG baseline wander.,” Med Biol Eng Comput **31** (5), 503-8 (1993).
[18]C Li, C Zheng, and C Tai, “Detection of ECG characteristic points using wavelet transforms.,” IEEE Trans Biomed Eng **42** (1), 21-8 (1995).

Appendix 1 Wavelet Filtering Method

A1.1 Aim

The initial aim of the authors work with wavelets was to pre-filter the signal before the modelisation stage of the project, in such a manner that the noise content of the signal was reduced, whilst all the interesting features of the waveform sequence remained intact. In normal high pass filtering the frequency response is distorted around the area of the band pass cut off limit, thereby distorting the signal. One of the principle features of wavelet analysis is this ability to pick out important features. Work was done specifically with orthogonal wavelets.

A1.2 Orthogonal Wavelets

It is possible with the Fourier Transform to decompose a signal into its frequencies, however all information regarding the time domain, i.e. when these frequencies occur, is lost. Wavelet analysis is designed to address these shortcomings, providing information of both frequency content and time content. The principle idea, as discussed before, is to generate a mother wavelet, which can be orthogonal or non-orthogonal. This wavelet is then translated and dilated to create a library of wavelets, each with a different frequency and a different position in time. By studying which of these wavelets best fit the signal, an idea is given as to the frequency content of the signal, as well as the position of these frequencies.

As with the non-orthogonal wavelets, the library is created via translation and dilation of a mother-wavelet, and represents a basis of the vector space. It is ensured that the basis is sparse i.e. all of the vectors in the library are linearly independent, by using the following means of dilation and translation:

$$W(a,b) = \varphi\left(\frac{x-b}{a}\right) \qquad \textit{Equation 7-1}$$

Where ϕ represents the orthogonal mother wavelet and $a = 2^{-j}$ and $b = k\,2^{j}$, where k and j are positive integers. The mother wavelet is created via tables, and cannot be represented analytically. As such there are only a limited number of orthogonal families of wavelets that can be created. Some of the more popular are Haar, Daubechies, Coiflet and Symmlet. These can be seen in figure 7-1. Within each family a further characteristic can be altered, that of the number of vanishing moments or par. [Vidakovic and Müller (1991)]

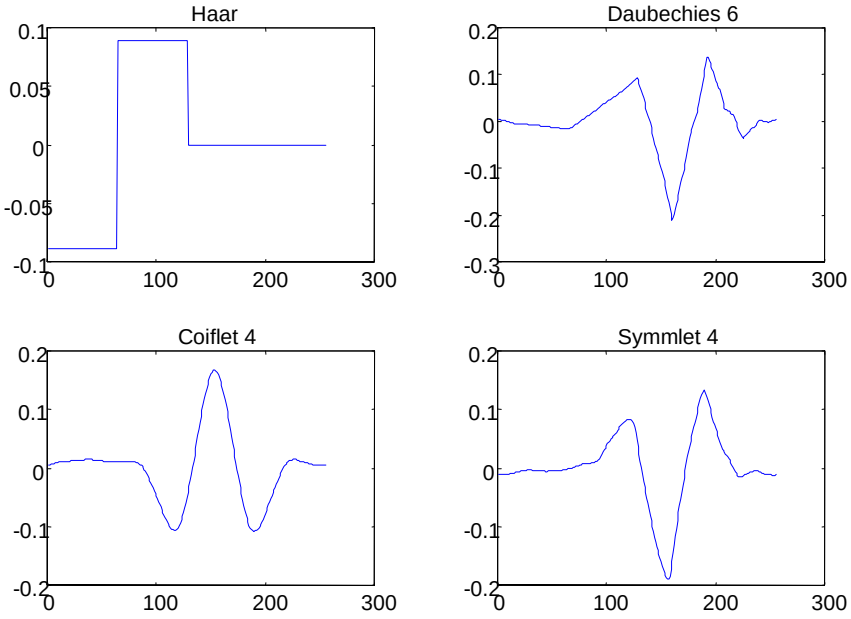

Figure 7-1: *Common types of orthogonal wavelets*

The great advantage with the use of orthogonal wavelets within this project, is that there is no need to undergo Gram Schmidt orthogonalisation to calculate the pertinence of each of the wavelets. The wavelets can simply be ordered by virtue of the angle their corresponding vector makes with that of the signal being studied (i.e. the scalar product). This decreases the speed of calculation dramatically.

An additional mathematical constraint of orthogonal wavelets, is that their length, (i.e. the number of points) is defined by 2^n, where n is an integer value. (Hence lengths of 64, 128, 256 etc, are all possible)

A1.3 Initial Filter

As discussed before, by adding all of the orthogonal wavelets in the basis, the signal can be reproduced exactly. However, this will have gained nothing as all of the noise will be modelled as well. If instead the wavelets are ordered by their angle with the signal, and then added, it is possible to eliminate as much or as little noise as is required, by stopping the addition at a certain point. Hence in the first 10 or so wavelets added, all of the major waveforms will be represented as well as any of the large variations in wavelet noise. After this smaller features will be added such as noise, activity around the isoelectric line, until the signal is completely represented at the end.

A1.3.1 Stopping Criteria

Work was then done on finding a means of automatically stopping the algorithm, once a desired accuracy was achieved. The method described here is commonly used in neural network design to create the best neurone architecture. The ECG signal is again split into windows, with the start and end falling at the points of the R waves. Points either side of the R

waves are then added, to make the vector up to a length possible for the orthogonal wavelets. In this case this was 1024 points.

This ECG vector **Y** is then split into two sets of points: a learning set $\mathbf{Y}_L$ comprising every second point of the window (hence creating a vector with 512 points), and a testing set $\mathbf{Y}_T$, comprising all of the other points. Wavelet analysis is then performed on the learning set $\mathbf{Y}_L$ using a library of orthogonal wavelets, represented by the matrix A:

$$A = \left| \boldsymbol{W}_1 \quad \boldsymbol{W}_2 \quad \boldsymbol{W}_3 \quad \ldots \quad \boldsymbol{W}_{512} \right| \qquad \textit{Equation 7-2}$$

where $\mathbf{W}_i$ represents each different wavelet in the basis. To order the wavelets, they are placed in order of their scalar products P_i. Hence:

$$P_i = \boldsymbol{W}_i^T \boldsymbol{Y}_L \qquad \textit{Equation 7-3}$$

Thus the wavelet with the largest value of P, will be the wavelet with the smallest angle to Y in the geometrical interpretation of the system (see chapter 3). The order of these wavelets is represented by a vector j, containing integer values, such that j(k) represents the wavelet number i, with the k^{th} largest scalar product. These wavelets are then added in this order of P to give a series of

$$\boldsymbol{Y}_M^N = \sum_{\kappa=1}^{N} P_{j(k)} \boldsymbol{W}_{j(k)} \qquad \textit{Equation 7-4}$$

where $\boldsymbol{Y}_M^N$ represents the sum of N ordered wavelets, (with integer values ranging from 1 to 512), chosen by the order in vector j. Thus for each value of N, we can calculate a different model of the learning points. As N increases, $\boldsymbol{Y}_M^N$ tends towards the vector $\boldsymbol{Y}_L$ until they are identical at N = 512. The aim is to find a value of N, such that all of the major features are modelled, but not the noise. This is achieved by calculating two errors for each value of N. These are the mean quadratic errors, and represent the average absolute difference between the wavelet sum $\boldsymbol{Y}_M^N$, and the learning set or the test set (these are termed MQEL and MQET).

$$MQEL^N = \frac{1}{512}\left(\boldsymbol{Y}_L - \boldsymbol{Y}^N{}_M\right)^T\left(\boldsymbol{Y}_L - \boldsymbol{Y}^N{}_M\right) \qquad \textit{Equation 7-5}$$

and

$$MQET^N = \frac{1}{512}\left(\boldsymbol{Y}_T - \boldsymbol{Y}^N{}_M\right)^T\left(\boldsymbol{Y}_T - \boldsymbol{Y}^N{}_M\right) \qquad \textit{Equation 7-6}$$

Thus as N increases MQEL will tend towards zero, though the value of MQET will tend towards decrease (as both the sets of points are similar), until a minimum is reached, where it will begin to rise again. This rise is due to the sum of the wavelets ‘overmodelling’ the points

in the learning set, and as N increases further, the wavelet sum will drift further away from these points. It is this minimum which will give an indication of the ideal number of wavelets needed to correctly model the signal, as it represents the point where the major features are modelled (and hence the best possible model for both sets of points), and the noise is not (i.e. when the model gets further away from that of the test points). This minimum can be represented on a graph of MQEL and MQEL against the number of wavelets added.

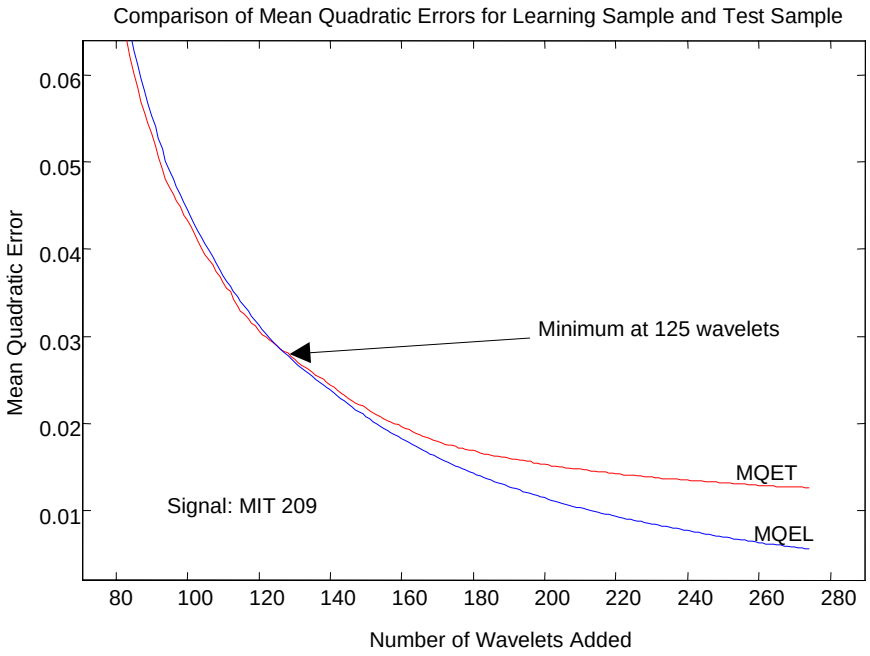

Figure 7-2: *Demonstration of no minimum in graph of MQET, with minimum in difference between curves shown in place*

It was found however, that in this type of modelling, the value of MQET never rises to the extent that a minimum can be seen precisely, with it levelling off to a fixed error value. (see figure 7-2). This was an unexpected result and was explained by the high sampling frequency of the ECG signal meaning that the learning set and test set, are very close in values.

A minimum was however seen in a comparison of the error values. This minimum corresponds to when the errors begin to separate, and the learning data set begins to be overmodelled. This is more clearly seen in a graph of the absolute difference between MQET and MQEL:

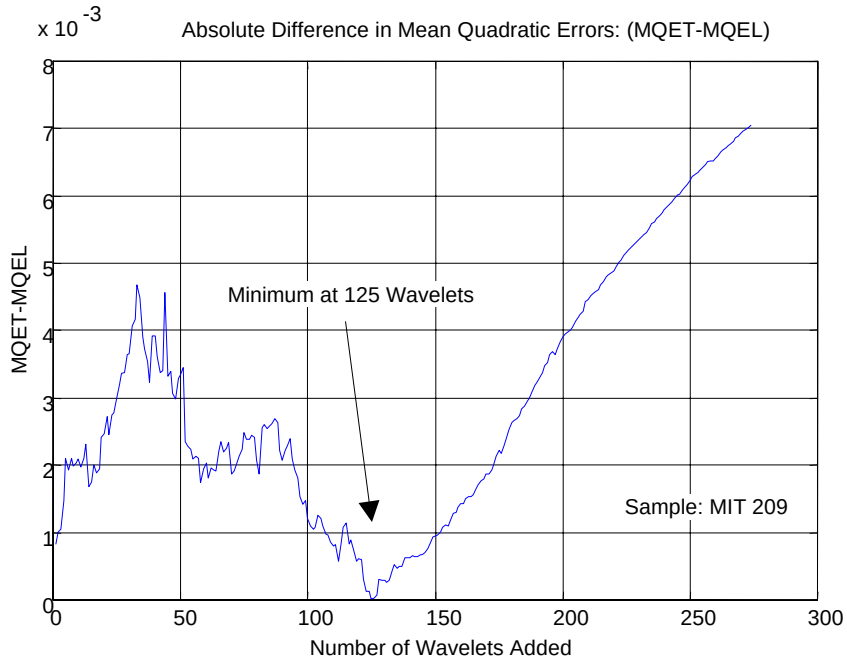

Figure 7-3: *Difference between error curves, demonstrating minimum at around 125 wavelets*

Qualitatively, this method was shown to give reasonable results in calculating a correct stopping point, though the fact that the MQET did not rise after passing that minimum, caused the author to look for other methods of stopping the addition (see figure 7-3). These included a probability study regarding whether a random vector is chosen before a wavelet, and if reducing the sampling frequency of the signal would effect the error calculation.

(NB. This research was never completed due to the choice of Gaussian wavelets in the modelling stage, and their accuracy in dealing with noise meaning that this variety of prefiltering was no longer necessary, especially given the large calculation time.)

It was noticed during this process, that during the addition wavelets noise in the isoelectric line was being modelled before important information on the P waves or T waves for example. Hence the greater the accuracy wanted on the PQRST waves (achieved by adding increasing numbers of wavelets), the greater the amount of noise that would be present in the final filtered signal.

A1.4 Spatial Filter

A system was developed to give greater precedence on wavelets modelling in the PQRST regions, rather than on the isoelectric line and hence noise. Several steps were involved:

1. Calculation of active zone
2. Create bias for each wavelet according to how well they fit into these active zones
3. Reordering the vector j, according to a bias from the scaling factors in 2
4. Adding the reordered wavelets to a required accuracy

A1.4.1 Calculation of Active Zone

It was this need to find the important regions of the signal, and hence the regions which should be modelled first, that lead to the development of the active zone algorithm, and subsequently the baseline and noise detection systems. The method used in this preliminary study was as outlined in section 4.2.1, taking a highly smoothing filter, to leave only the R waves and information surrounding them. This filter was based on the assumption that all of the interesting waveforms were to be found around the R waves.

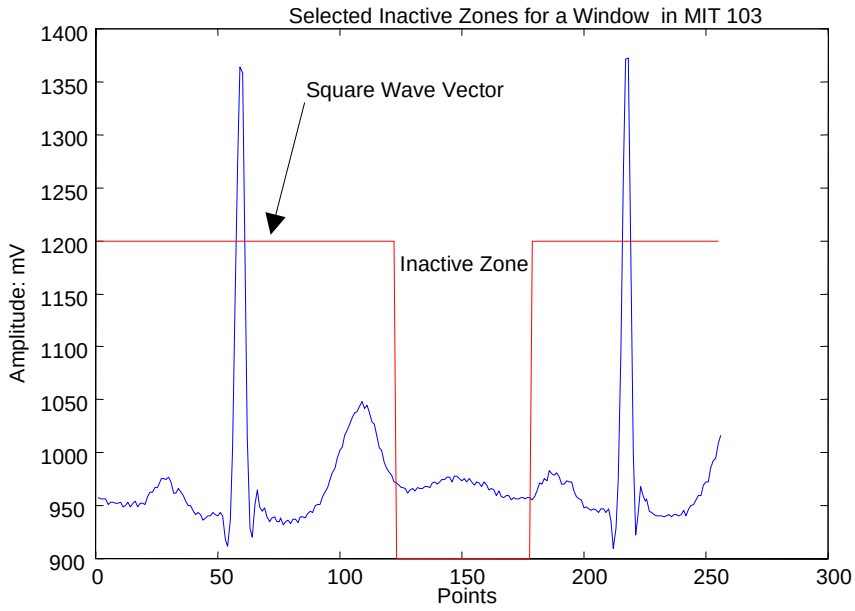

Figure 7-4: *ECG signal to be tested, shown with the chosen inactive zone*

This then gave a square wave over the ECG window that can be seen in figure 7-4, with the square wave vector taking values of either 0 or 1, referring to inactive or active points respectively. This active zone vector is termed **Z**. For the purpose of this example an orthogonal Daubechies wavelet basis (par 6) with 256 points was used, thus the ECG window **Y**, has also 256 points.

A1.4.2 Calculation of Bias

In order to influence the selection order of the orthogonal wavelets, it was necessary to adjust the value of the scalar products, P_i. To do this a bias was given to each value of P_i, determined by how well each wavelet fitted within the active zone vector. By applying the bias, a wavelet with the majority of its signal in the inactive zone will be made less pertinent than one in the inactive zone. This bias was calculated from the scalar product of the active zone vector and each wavelet.

$$B_i = \boldsymbol{W}_i^T \boldsymbol{Z} \qquad \textit{Equation 7-7}$$

where i = 1, 2, … 256 (as there are 256 wavelets in the basis).
These bias values are demonstrated in figure 7-5.

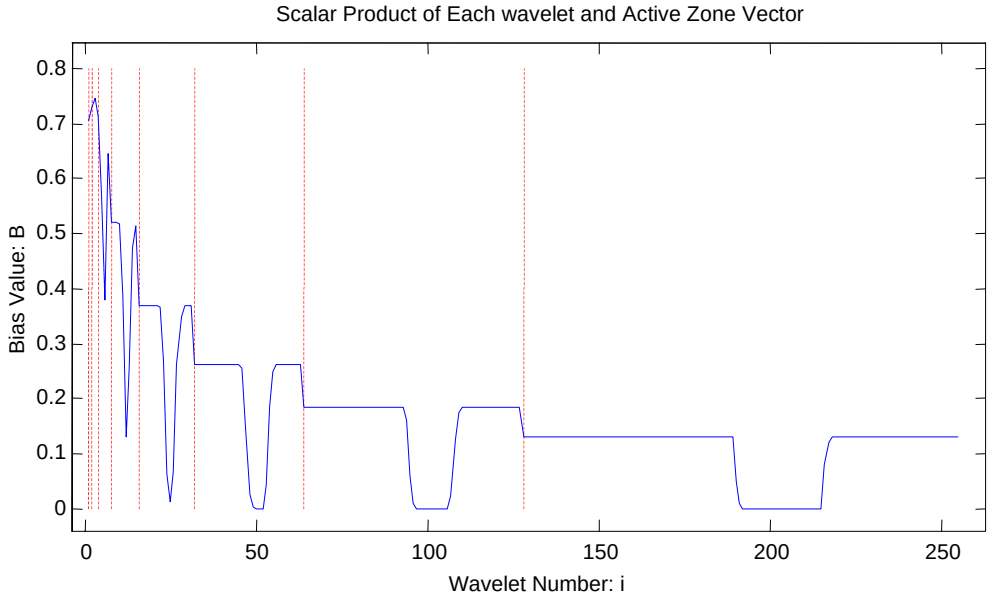

Figure 7-5: *Demonstration of dilation levels within wavelet basis, showing troughs for wavelets that correspond to the inactive zone*

Figure 7-5, demonstrates clearly the 8 different levels of dilation that are present in the wavelet library. For each dilation level, there are an increasing number of translations, given by equation 7-1. Hence troughs can be seen in the bias values. These troughs correspond to the translations of wavelets in each dilation level that correspond to the inactive zone. Thus we have achieved the aim of placing a bias on those wavelets not within the active zones. These bias values were then scaled between 0 and 1, such that the dilation level, does not affect the bias given to each wavelet (see figure 7-6).

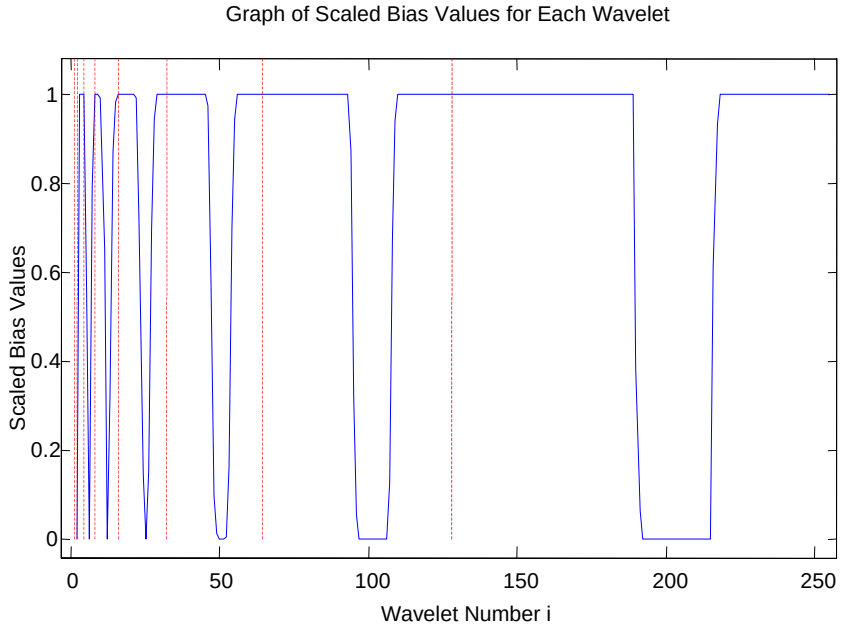

Figure 7-6: *Bias values scaled for each dilation level*

The original scalar products, of the signal and the individual wavelets, are then multiplied by these new bias values and then placed in order. The same process as in the initial filtering

method is then applied, adding wavelets in order until the stopping criteria is reached and a desired accuracy has been obtained.

A1.4.3 Results

Although the results for this filtering method were never fully researched, the following qualitative results can be shown. The following figures (7-7 and 7-8), demonstrate the different speeds with which noise is modelled in two cases, ordered using projections with the described bias, and the projection values without the bias.

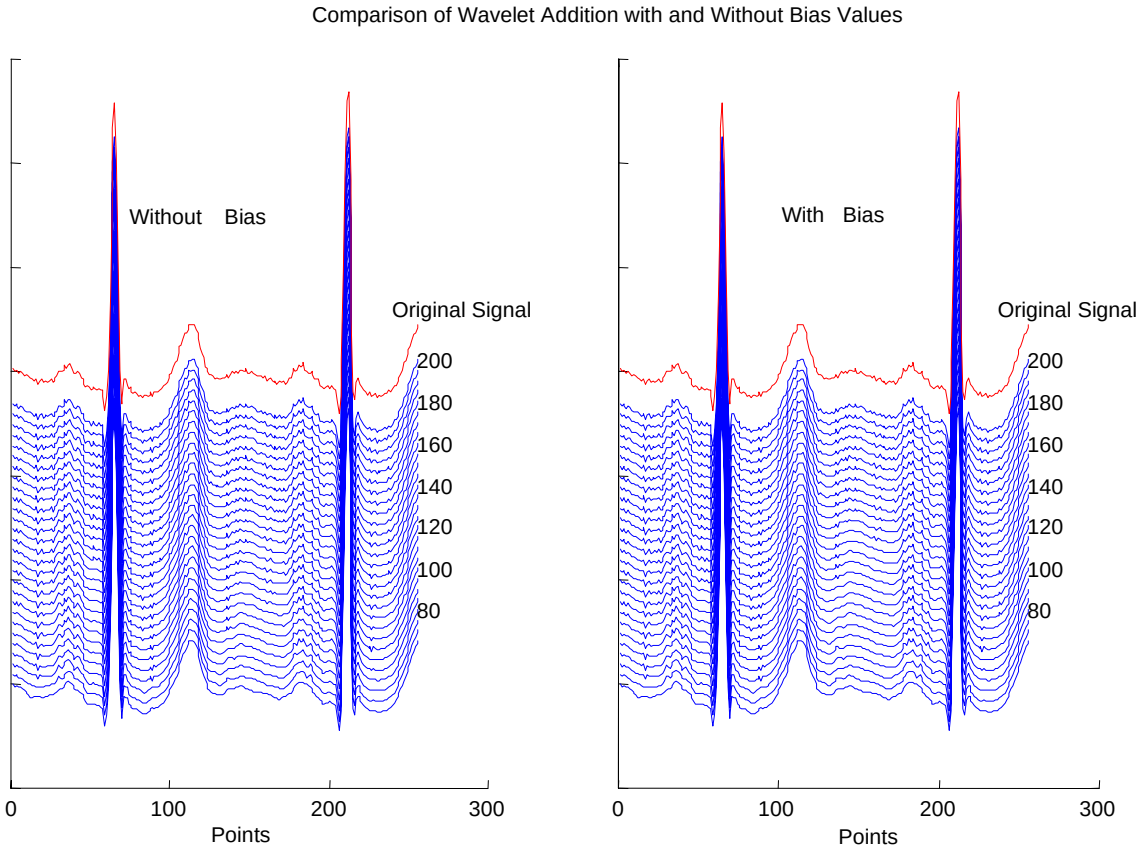

Figure 7-7: *Comparison of sums of wavelets with and without the applied bias*

As can be seen in the graph, in the inactive region between the T and the P wave, the noise is shown to be reproduced much slower (more wavelets needing to be added before it is modelled; 110 with bias, 80 without) in the case with the bias, than without. Additional features are that although a bias reduces the chance of a wavelet corresponding to the inactive zone being selected, all of the major characteristics of this area have been represented. This will mean that fewer numbers of wavelets are required to recreate the signal to the desired accuracy on the important waveforms, and thus reduce the overall amount of noise in the window.

This noise reduction can be seen more clearly in figure 7-8.

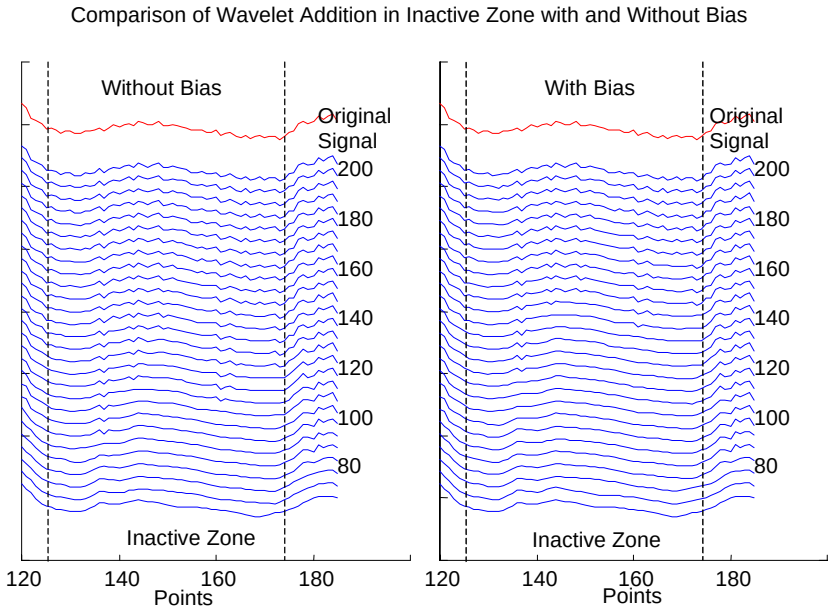

Figure 7-8: *Zoom on inactive zone, showing dramatic improvements on noise within bias zone. Noise is modelled some 40 wavelets later in the signal with bias*

In summary, the results show that by applying the bias to the wavelets, it is possible to influence the order of wavelet selection, and hence reduce the amounts of noise in specific areas of the ECG signal.

This research was never completed due to the choice of Gaussian wavelets in the modelling stage, and their accuracy in dealing with noise meaning that this variety of prefiltering was no longer necessary, especially given the large calculation time when the bias is considered.

Appendix 2. Additional Results

A2.1 Baseline Algorithm Results

These results represent examples from the MIT database that demonstrate the baseline detection algorithm, correctly fulfilling both aims of isoelectric line detection, and baseline wandering removal. These results are separated by difficulty type: i.e. difficult pathologies, high frequency noise, low frequency noise.

A2.1.1 Signal Type

Normal Signal

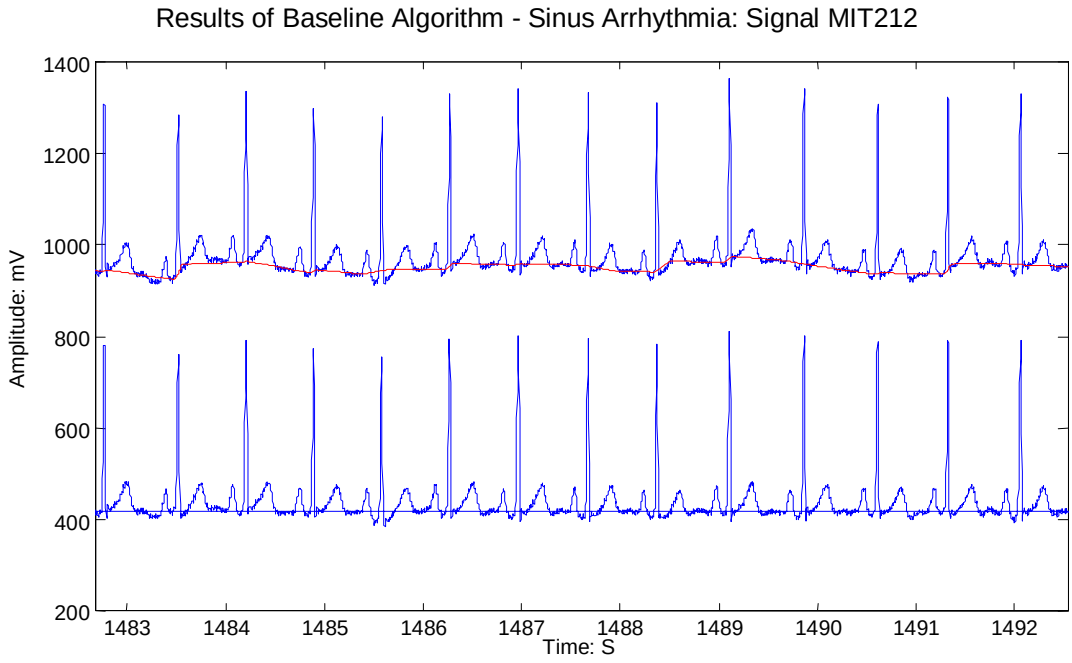

Figure 8-1: *Correct baseline detection for normal ECG*

This graph demonstrates the algorithm's ability to interpret all areas of the isoelectric line, and presents no waveform deformation.

Premature Ventricular Contraction

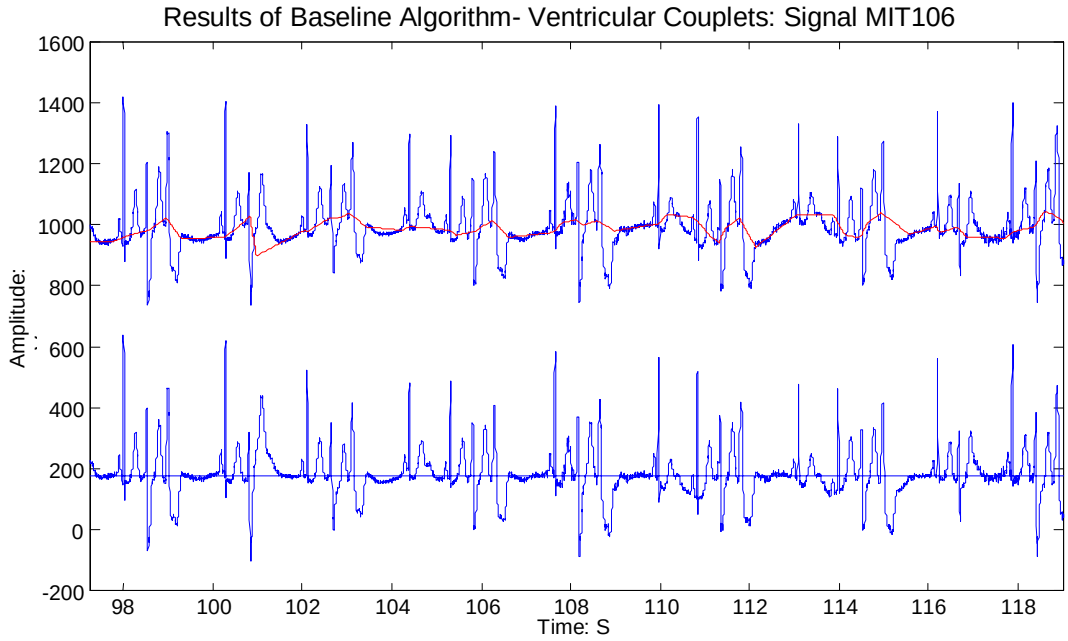

Figure 8-2: *Correct baseline detection for signal with large variations in pathology size and shape*

This graph was included to show the ability of the algorithm to work through signals with great complexity, correctly adapting to the low frequency noise at the same time.

Ventricular Tachycardia

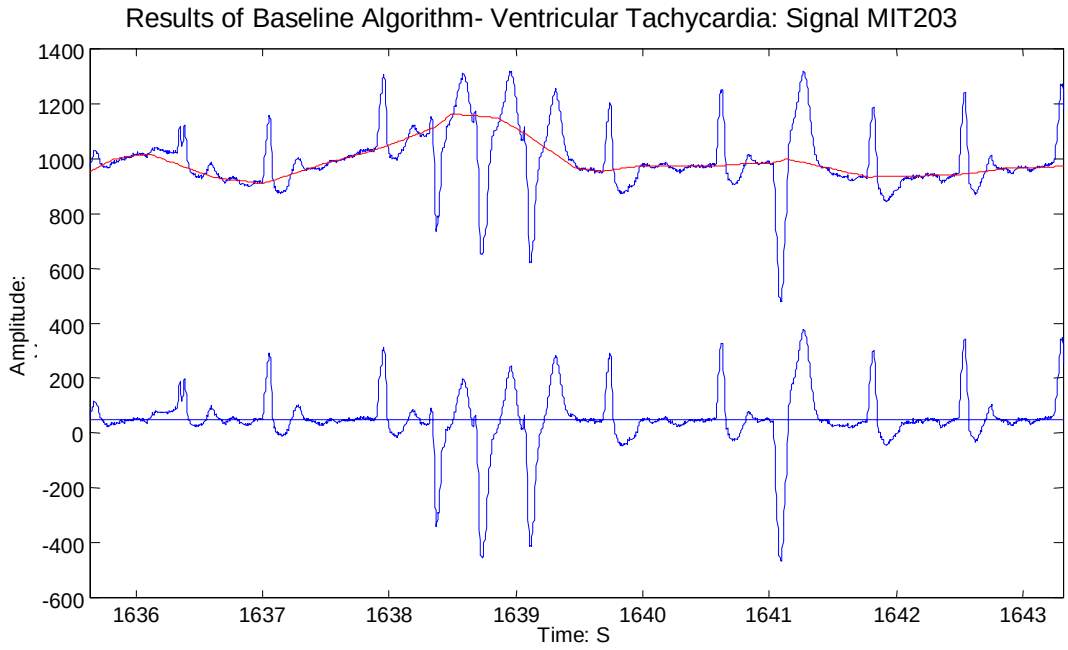

Figure 8-3: *Correct baseline detection for signal with limited isoelectric regions, and some baseline wandering*

Again, this is an example of a very difficult signal, owing to the lack of well-defined isoelectric zones, as well as the additional low frequency noise. However, the algorithm has adapted well to the different pathology presented, and a correct baseline has been given.

Pace Maker

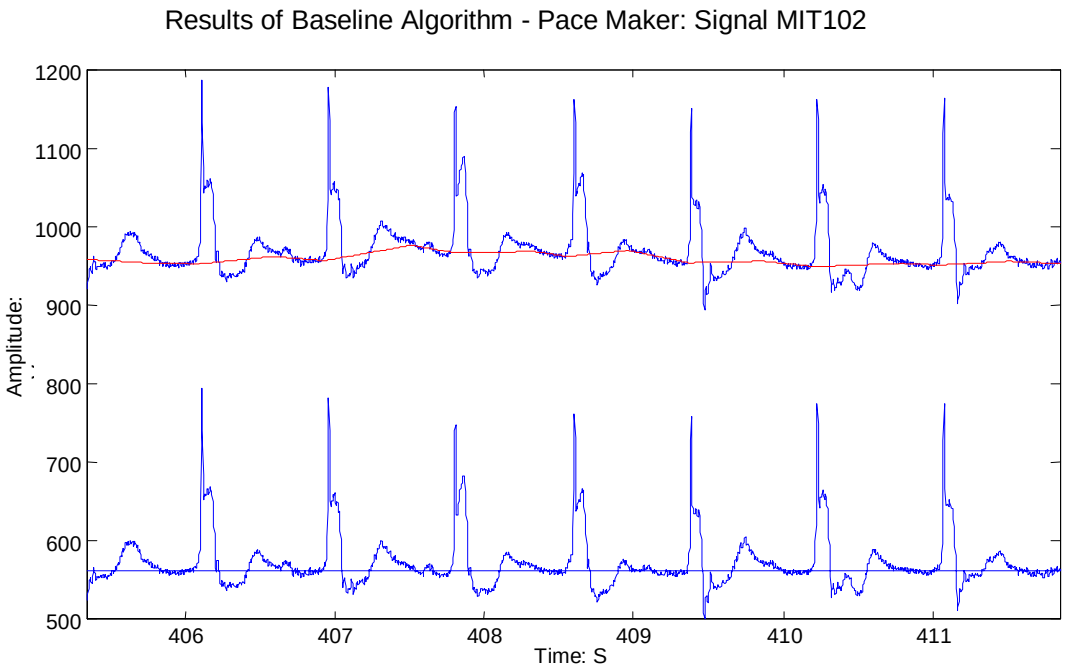

Figure 8-4: *Correct baseline detection for ECG signal with artificial pacemaker*

This result again demonstrates the algorithm working in an area of limited and deceptive baseline. Although there is only the P-R segment not active in this signal, it has managed to pick this region, and not that of the slowly reducing T wave.

A2.1.2 High Frequency Noise

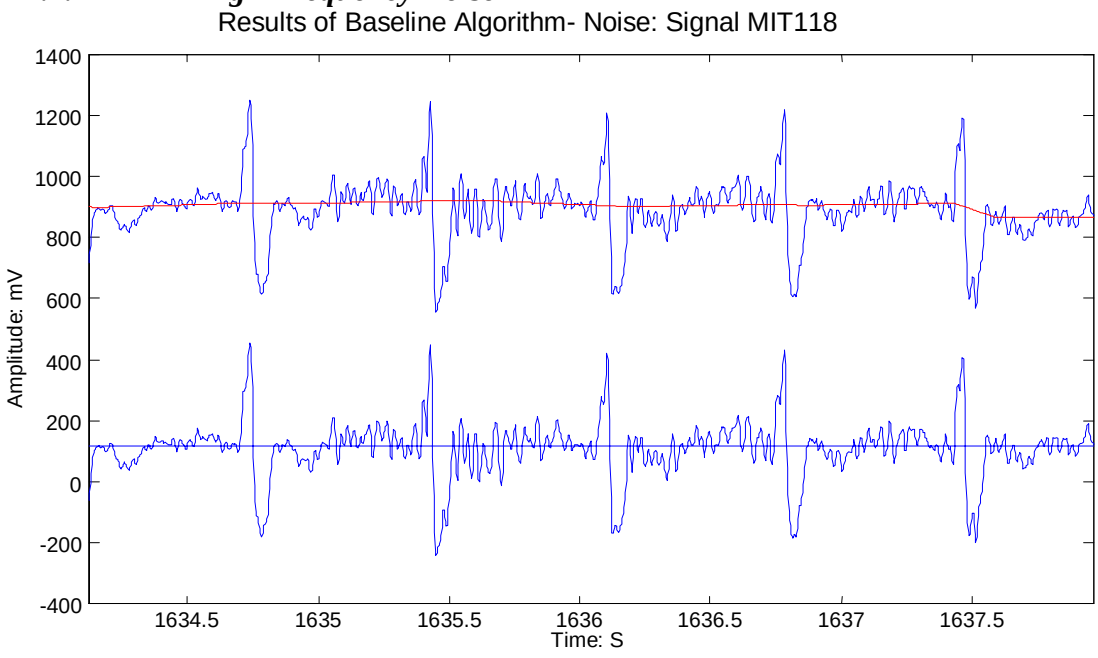

Figure 8-5: *Correct baseline detection through high frequency noise*

As can be seen in the graph, the system has adapted well to the varying amplitudes of high frequency noise and gives a very stable result. This is principally due to the use of the

Gaussian filter to detect the active zones, a method that is able to ignore the noise, and look at the interesting signal beneath.

A2.1.3 Low Frequency Noise

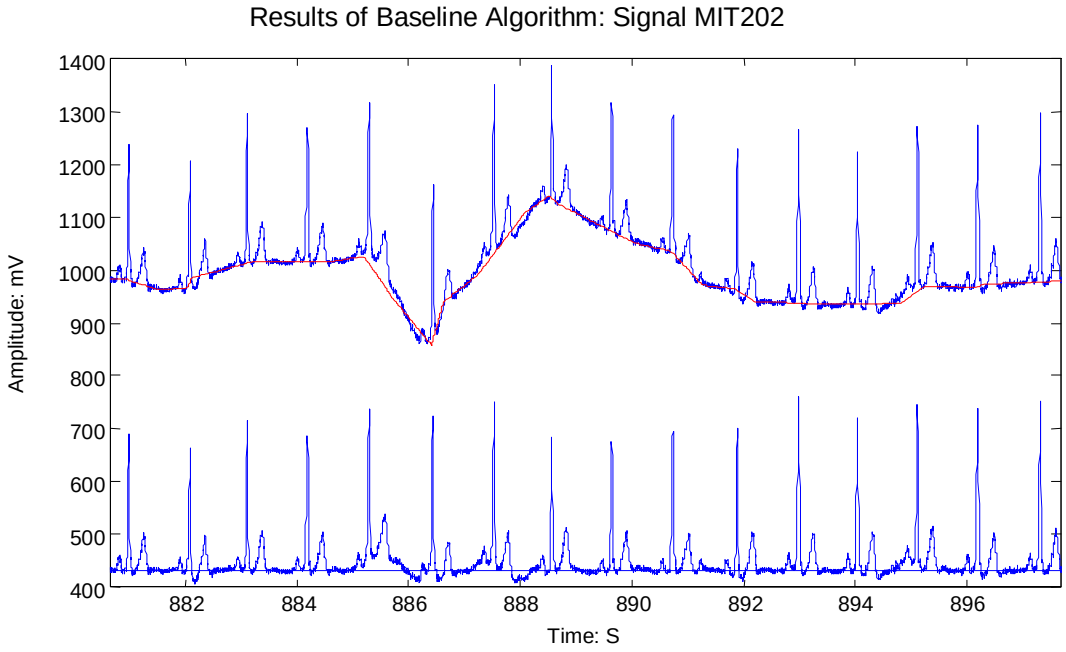

Figure 8-6: *Correct baseline analysis of signal with sharp baseline wandering*

Low frequency noise is the principle enemy of this algorithm, but as will be seen later, this is only in specific circumstances. In the majority of cases, as is seen in the graph, the algorithm is able to adapt to even the most violent changes, and still give a usable baseline.

In summary, these results demonstrate the accurate removal of baseline wandering for a great majority of the signal.

A2.2 Comparison of accuracy with PR isoelectric line

These results are designed to compare the isoelectric line calculated by the proposed algorithm and that by taking a point 40mS before the R Wave (as discussed in section 5-4). All of the results demonstrate a more accurate baseline when using the proposed algorithm, and various reasons for this are displayed below. These figures represent signals from the MIT after baseline wandering has been removed to some degree via a high pass filter. Thus the results are representative of the isoelectric line detection, rather than baseline wandering removal and a flat line interpolation is presented (as described in section 5.61). In each figure, 3 baselines are presented, and the baseline discussed in each section refers to that of the central.

A2.2.1 Difficult Pathology

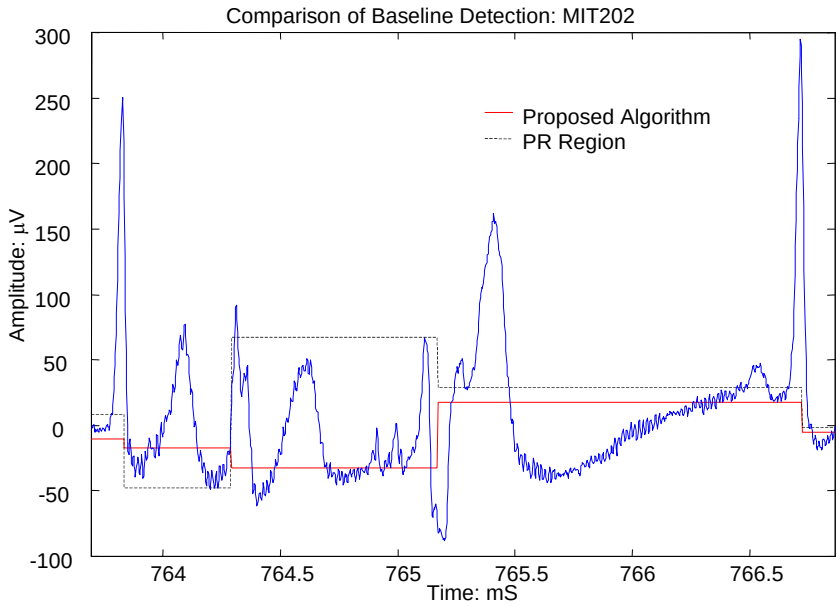

Figure 8-7 : *PR baseline failure due to large and inverted R wave*

Figure 8-7, demonstrates the higher accuracy of the proposed algorithm, being able to look for other isoelectric regions, rather than those just in the PR region. This is also an example for the need for perfect R wave detection.

A2.2.2 Low Frequency Noise

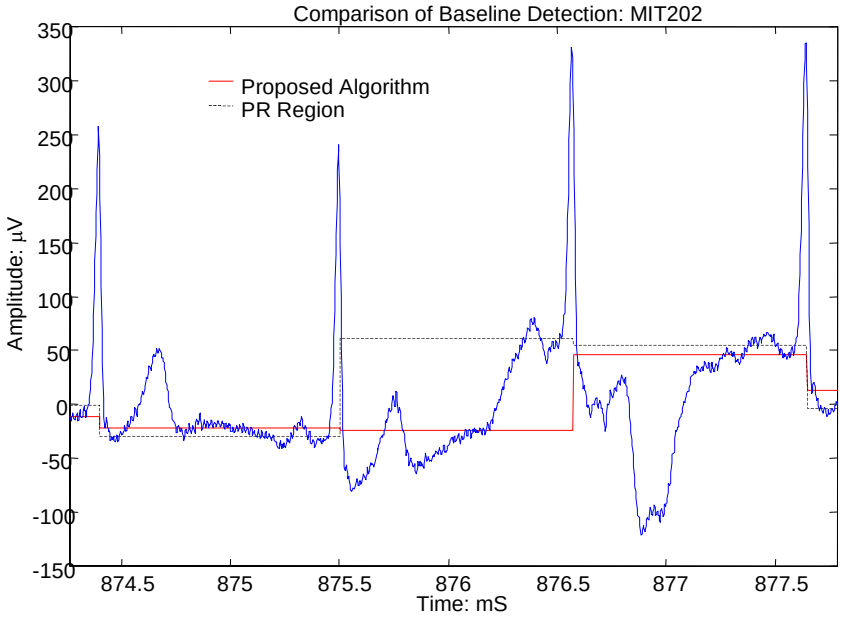

Figure 8-8 : *PR baseline error due to residual low frequency noise, raising the PR segment*

Figure 8-8, again demonstrates higher accuracy from the baseline algorithm, with the PR baseline unable to account for the sudden variation in amplitude. As is seen, the T wave is completely lost by the PR baseline.

A2.2.3 Inverted R wave

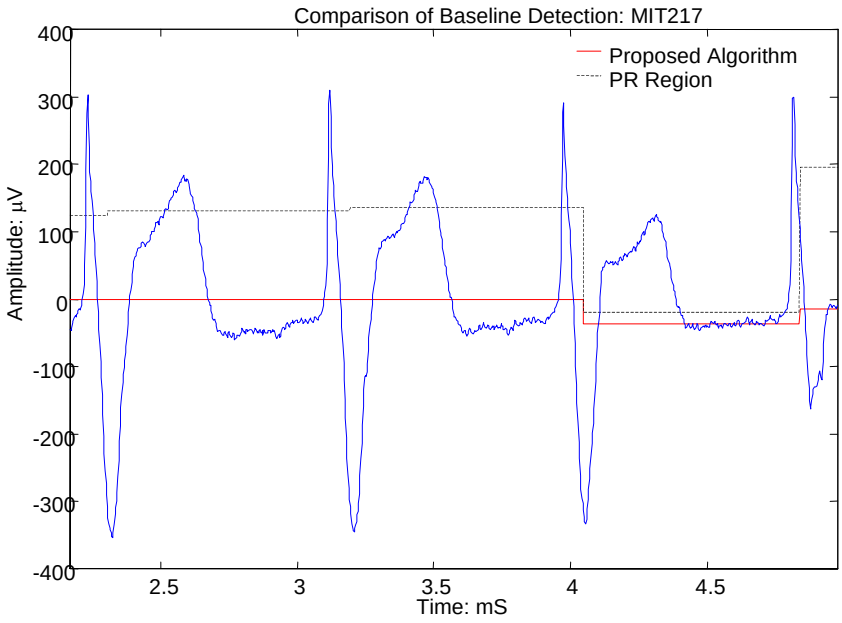

Figure 8-9 : *PR baseline error due to inverted R wave*

In figure 8-9, the R wave has been inverted, and as such, a point 40ms before the R wave will give a widely inaccurate baseline. In this, the baseline chosen does not pass through any isoelectric regions.

A2.2.4 Wide R Wave

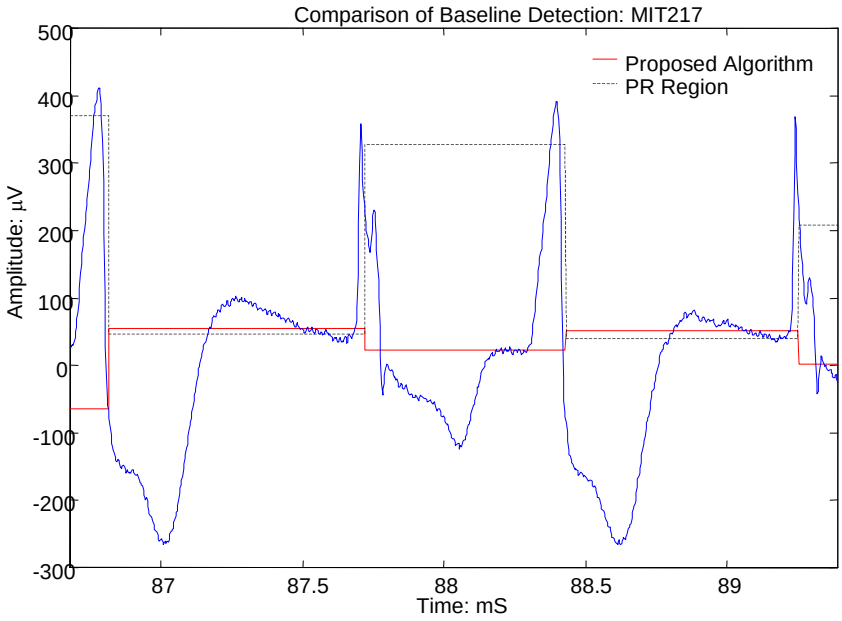

Figure 8-10 : *Poor PR baseline due to wide R wave*

In figure 8-10, the R wave has been enlarged by a ventricular extrasystole, and thus the 40ms point chosen again falls on a point in the R wave, causing a widely inaccurate result.

A2.2.5 Raised PR Region

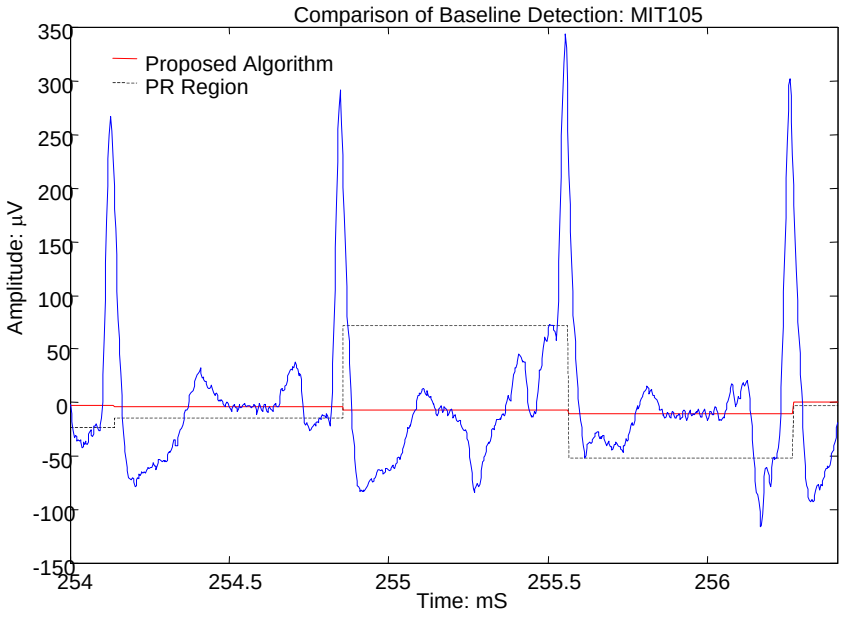

Figure 8-11 : *PR baseline error due to a raised PR segment*

This result, again demonstrates a raised PR segment, though this is due to the pathology type rather than any baseline wander.

A2.2.6 40mS point Falling on R wave

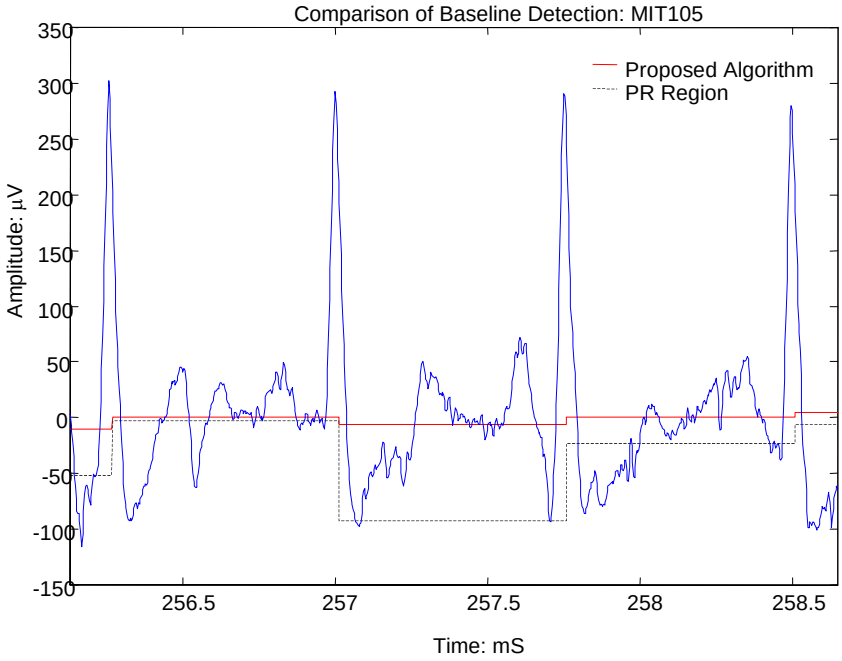

Figure 8-12: *PR baseline error due to large Q wave*

This is a common error in the PR baseline, and is unavoidable due to the lack of any PR isoelectric region. As is seen in figure 8-12, the point 40ms before the R wave, falls on the minimum of the Q wave, thus producing this very inaccurate result.

A2.2.7 Summary

The results demonstrate that the PR baseline is too unreliable to be used as an estimate of the isoelectric line, when the accuracy needed is taken into account. Large errors are caused by the systems over reliance on precise R wave detection, as well as the actual presence of a PR segment. It is also evident from the figures, that in all of these highlighted cases, the proposed baseline algorithm has been able to predict the baseline to a much higher accuracy.

A2.3 Noise Estimation in MIT Signal Database

The following figures are included so as to demonstrate qualitatively the results using an non-artificial ECG signal. The signal used is number 105 from the MIT signal database.

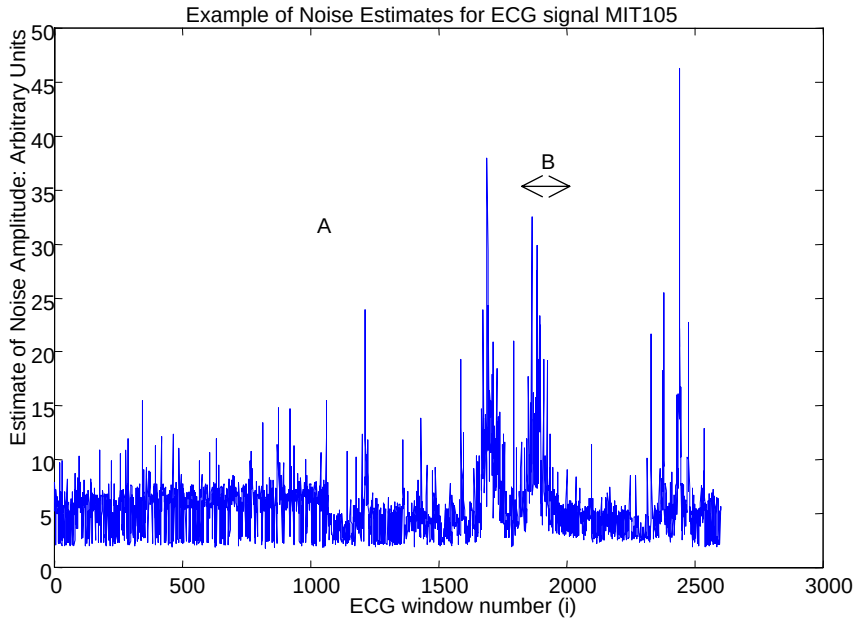

Figure 8-13: *Noise estimation in MIT 105, demonstrating several large peaks of non-analysable high frequency noise*

Figure 8-13, demonstrates several large peaks in noise estimation. It is these values that must be thresholded, thus determining which of the ECG windows do not pass to the modelling stage of the group's work. Several examples of these signals are shown in the next few figures.(feature A and features B).

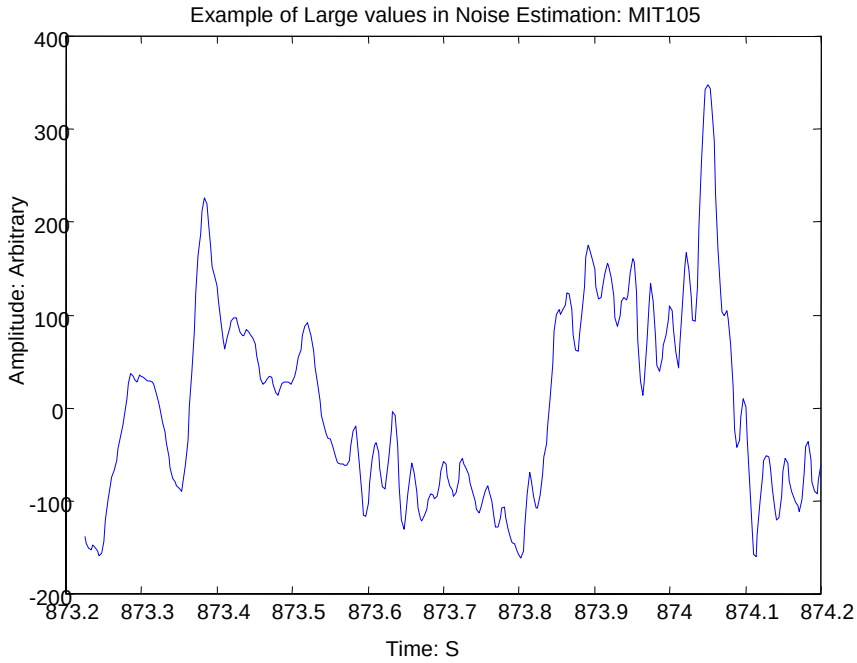

Figure 8-14: *Signal from feature A in figure 8-13*

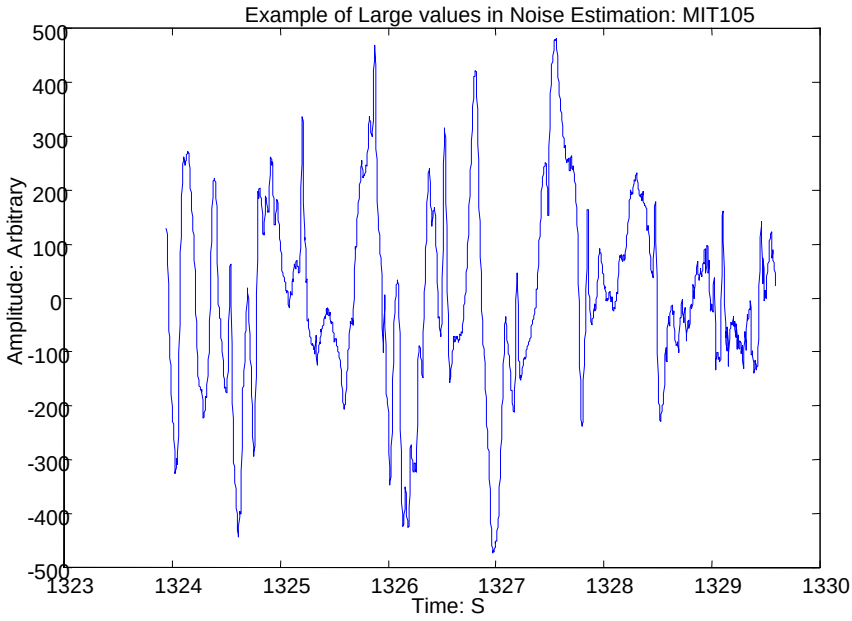

Figure 8-15: *Signal from feature B in figure 8-13*

These figures are simply to demonstrate some of the areas of the ECG signal, that would be removed using a threshold on the estimated values of the noise.

Appendix 3. Additional Figures

A3.1 Example of 12 Lead ECG Signal

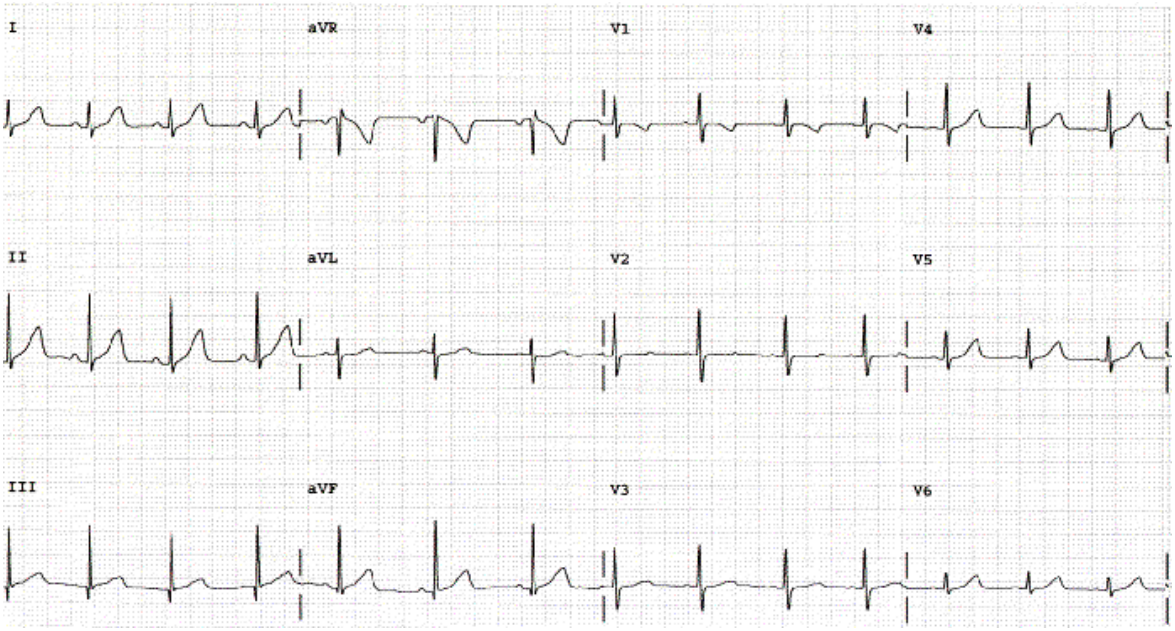

Figure 9-1: *Diagram of standard 12 lead ECG signal, with normal heartbeat*

Figure 9-1, This demonstrates the different variations in ECG signal as lead changes. Each of the 12 signals above represents the same signal, though each waveform is accentuated differently depending on the axis of the lead.

A3.2 Examples From the Simulator

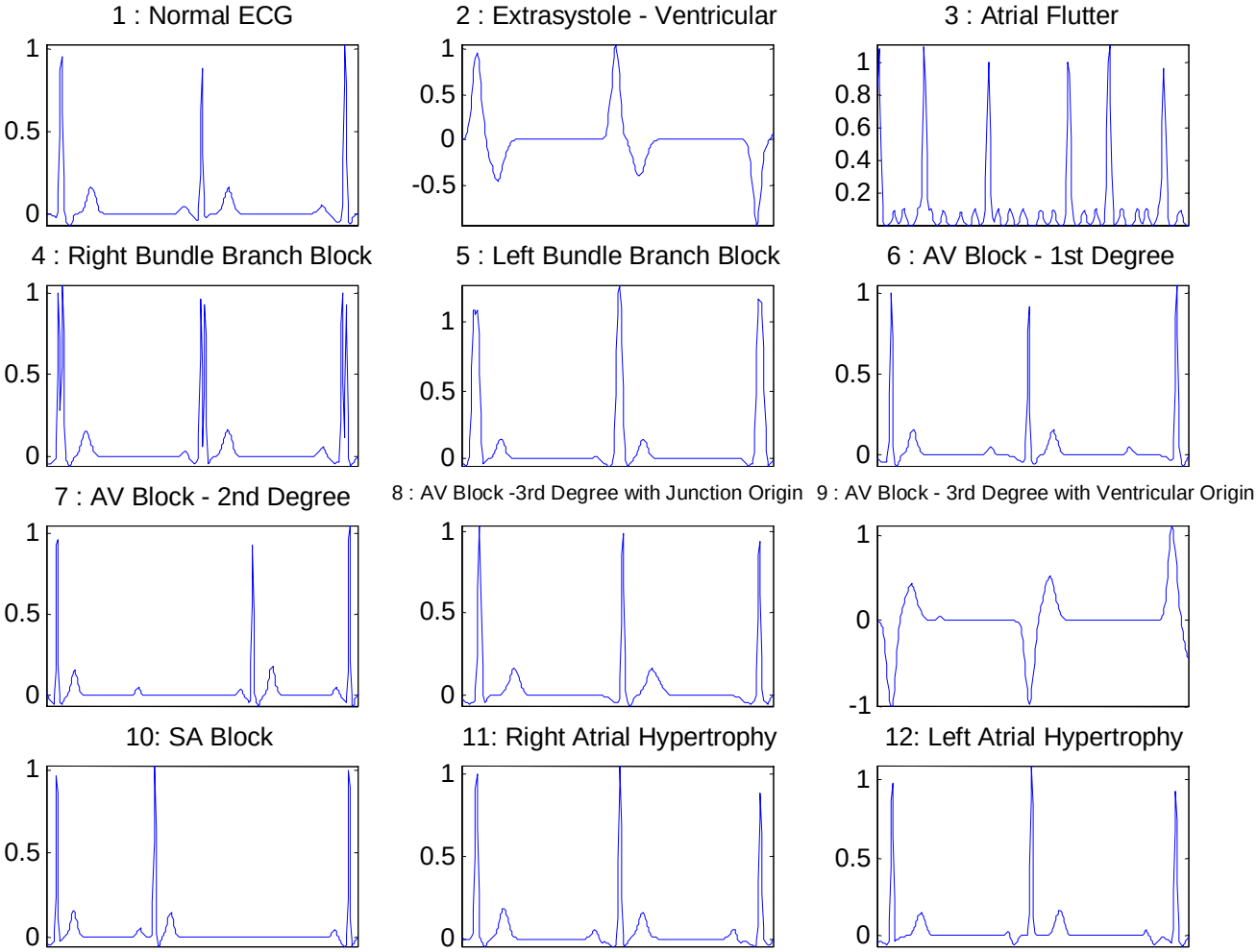

Figure 9-2: *Diagram of all different pathology types that can be created with the ECG simulator as designed by Remi Dubois of ESPCI*

This figure is designed to give an indication of the different types of signal that were created with the simulator. In addition to pathology type, it was possible to vary the heartbeat freauency in the simulator. The simulator is designed to be as close as possible to a normal heartbeat, and as such there are numerous random elements introduced, such that each heartbeat is never identical to the next. As can be seen in figure 7-2, the signals are perfect, and have no noise added. This allowed the author to have full control over the low and high frequency noise that was added to the signals, over the course of the project. These signal type numbers are referred to within the results sections of the report.

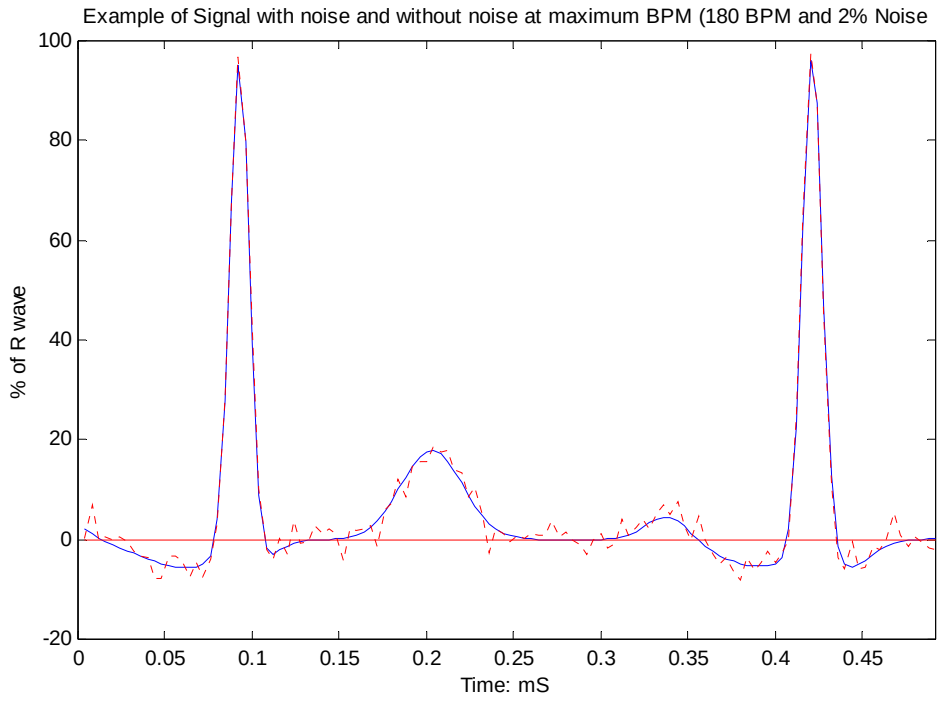

Figure 9-3: *Example of simulated signal at maximum heartbeat frequency -180bpm*

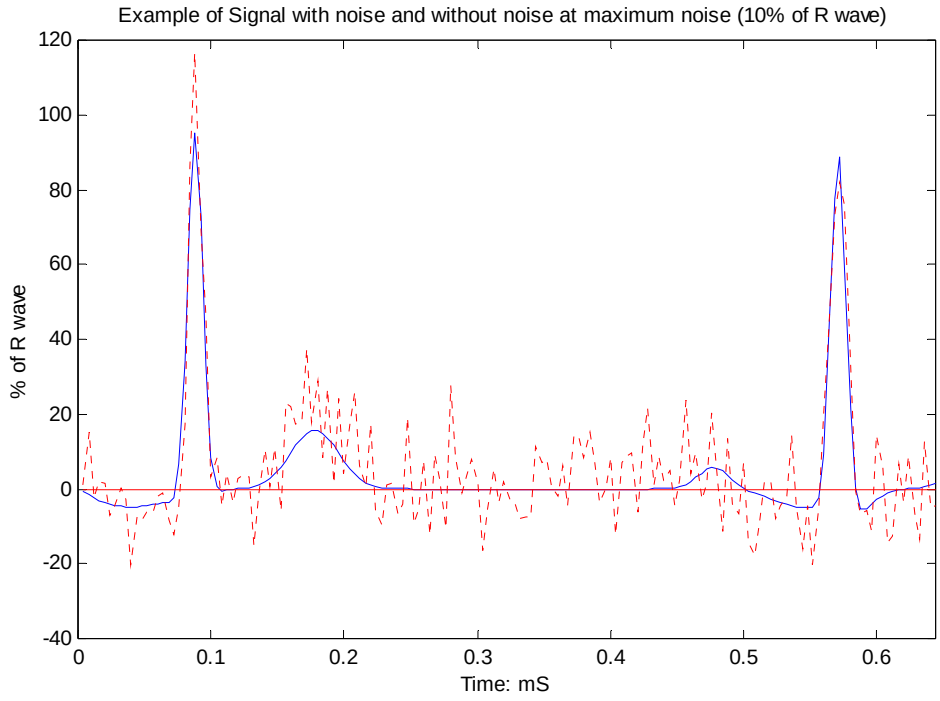

Figure 9-4: Example of simulated signal with 10% of noise added

A3.3 Examples from ELA Test Signal

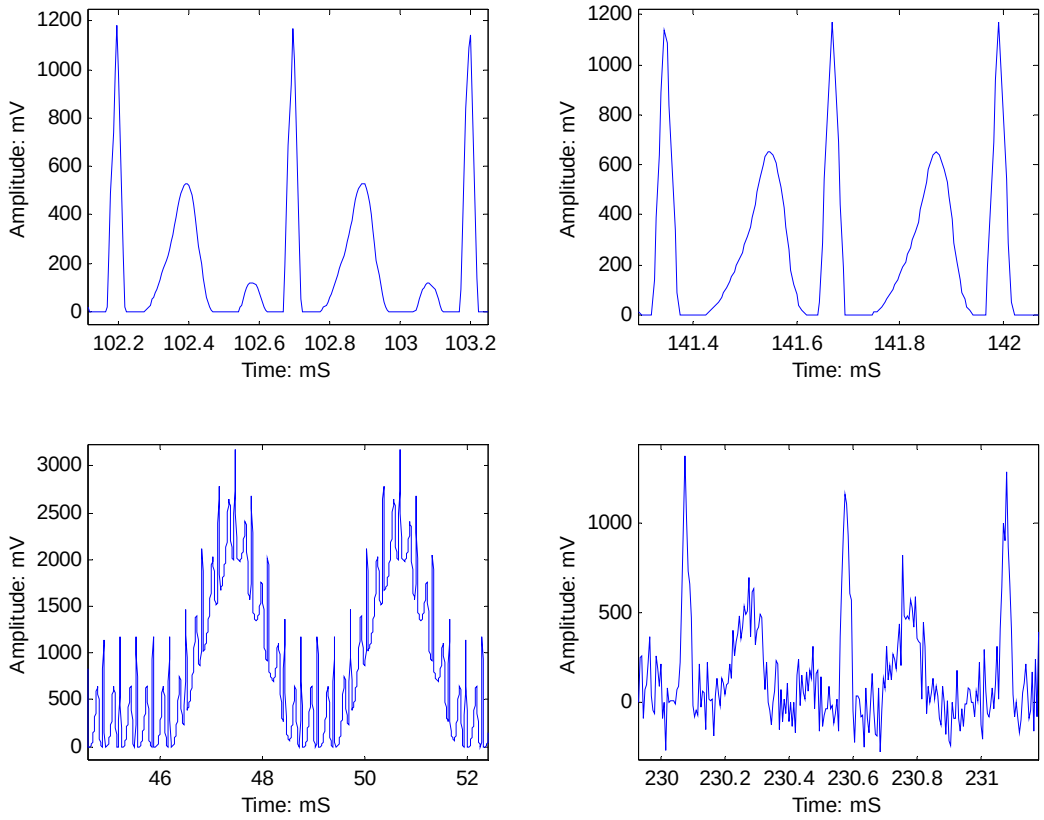

Figure 9-5: *Examples of ELA test signal: top left – Heartbeat of infant under 1 month (large waveforms)*

Top right- rapid heartbeat frequency (limited isoelectric line)

Bottom left- sharply rising baseline wander

Bottom right – High frequency noise added to signal

The ELA test signal consists of around 1000 heartbeats, and is graded with three different amplitudes of high frequency noise. The pathology type is particularly difficult due to the large T wave, in comparison to the R wave. Additional areas of sharp baseline wander are visible in all three levels of noise.

Appendix 4. Gram Schmidt Orthogonalisation – Matlab Code

This program relates to chapter 3 in the main report, and describes a Matlab program to select the most pertinent wavelets from a library, to model an ECG signal.

```
function y = sortGS2(W,A,num);
%Order num vectors in base A, via gram schmidt orthogonalisation, %with
vector W
% W:   The Vector being analysed
% A:   The wavelet library
% B:     Matrice orthonormal created during program
% num: The number of wavelets to be chosen from library A

% Author:  JS'00

%++++++++++++++++++++++++++++++++++++++++
%Arguments
if nargin==2
   num = size(A,1);
end

%++++++++++++++++++++++++++++++++++++++++
%Normalise base such that all vectors have lenghth 1
for i=1:size(A,1)
   A(i,:)=A(i,:)/norm(A(i,:));
end;
%++++++++++++++++++++++++++++++++++++++++
%Setup initial variables
[numvec,dim] = size(A);
A1 = A;
A2 = A1;
W1 = W;
B(1,dim) = zeros;

%++++++++++++++++++++++++++++++++++++++++
% Loop until num wavelets have been chosen from the base
for n = 1:num
   %++++++++++++++++++++++++++++++++++++++++
   %Calculate maximum projection indmax and hence closest wavelet to W
   proj = (W1 * A1')/ norm(W1);
   [Z,indmax] = max (abs(proj (n:numvec)));

   %++++++++++++++++++++++++++++++++++++++++
   %Place Wavelet base in order
   indmax = indmax+n-1;
   order(n) = indmax;
   A1(n,:) = A2(indmax,:);
   A1(indmax,:) = A2(n,:);

   %++++++++++++++++++++++++++++++++++++++++
   % Calculate matrix B, orthogonal to chosen wavelets
   B = Gschmidt(A1,B,n);

   %++++++++++++++++++++++++++++++++++++++++
```

```
   %Calculate projection of vector W in orthogonal subspace
   Wpara = 0;
   for p = 1:n
      Wpara = Wpara + (W1*B(p,:)') * B(p,:);
   end
   Wperp = W1 - Wpara;
   W1 = Wperp;

   %++++++++++++++++++++++++++++++++++++++++
   %Calculate projection de matrix A in subspace
   for j = n+1:numvec
      Apara = 0;
      Apara = Apara + (A1(j,:) * B(n,:)') * B(n,:);
      Aperp= A1(j,:) - Apara ;
      A1(j,:) = Aperp;%./ ( norm(Aperp) + (norm(Aperp)==0) );
   end
   A2 = A1;
   if mod(n,10) ==0, disp(n);end
end

%++++++++++++++++++++++++++++++++++++++++
Place wavelets in order for output
j = 1;
refer1 = 1:size(A,1);
refer2 = refer1;
while j <num+1;
   i = order(j);
   refer1(j) = refer2(i);
   refer1(i) = refer2(j);
   refOut(j) = refer1(j);
   refer2 = refer1;
   j = j+1;
end
y = refOut;
%====================================
% refout contains the index numbers of the num most pertinent wavlets in
the base



%************************************************************************
************
function y = Gschmidt(A,B,n);
%Calcule matrice B, le base orthonormal a wavelet choisi
p = 1;
B(n,:) =A(n,:);
while p < n
   B(n,:) = B(n,:) - (B(p,:)*A(n,:)')/(B(p,:)*B(p,:)')*B(p,:);
   p = p + 1;
end
B(n,:) = B(n,:)/norm (B(n,:));
y = B;
```