**中期研发报告**

1. **内容与概述**

在太一给出的脉搏数据的基础上，完成对数据进行去噪和特征提取处理：

1. 脉搏数据信号噪声处理：信号噪声处理用于提供更稳定合理的数据，分为如下三个算法模块：

a. 采用信号处理技术去除脉搏数据信号的高频和低频噪声；

b. 采用信号处理技术去除脉搏数据信号的基线漂移；

c. 采用信号处理技术去除外部压力不一致导致的绝对值偏移噪声，并使信号归一化。

1. 脉搏数据信号特征提取：特征提取用于提取各类脉搏信号的特征值，分为如下几个算法模块：

对于单个独立的测量点提取：

a. 峰值/谷值；

b. 周期/脉率；

c.节律；

对于多个测量点联合提取：

d. 脉宽；

分析脉搏数据信号在给出外压值改变下的动态响应特征：

e.出现脉波的脉压峰/谷值和外压值

1. **开发环境**

|  |  |
| --- | --- |
| 开发环境 | 版本号 |
| Python | 2.7.10 |
| Numpy | 1.8.0 |
| Scipy | 0.13.0 |
| Matplotlib | 1.3.1 |

1. **数据定义**
2. **采用数据说明**

目前数据包括3份：

1. 太一第一次给的测试数据;
2. 140个病人的数据;
3. 清华自己测的数据，使用梳子牙刷进行辅助。
4. **输入输出定义**

输入：

含有寸关尺三部共计30个采样点的波形数据。具体采样点分布如下：

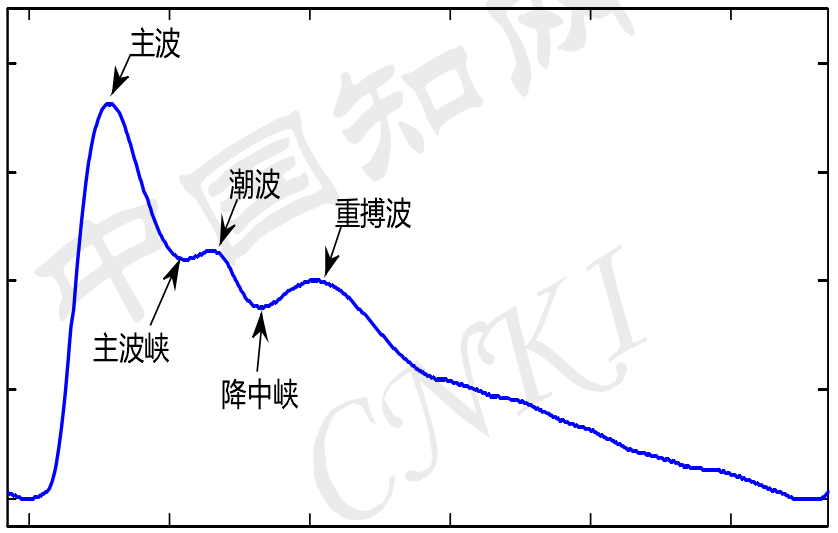
| 部位 | 采样点数 |
| --- | --- |
| 寸 | 两行12个点，每行6个点 |
| 关 | 两行12个点，每行6个点 |
| 尺 | 一行6个点 |

输出：

输出30个采样点的特质值，包括：

| 特征值 | 描述 |
| --- | --- |
| 幅值 | 由于外压会影响幅值，并且无法获得外压。因此给出每一个点的所有幅值 |
| 脉宽 | 每一个周期的数据，如果该行数据可以以二次曲线拟合，则给出脉宽，如果无法拟合，则脉宽为0 |
| 周期 | 周期对于所有采样点相同，因此对30个采样点的所有波形周期取平均，作为最终这份数据的周期值 |
| 波形细节-(H1~H6, T1~T6) | 如果去噪后波形噪声低，则可以取出波形细节，共有（H1,T1）,（H2,T2）,（H3,T3）,（H4,T4）,（H5,T5）,（H6,T6）6个特征点，12个特征值。  如果去噪后波形噪声高，则所有波形细节特质值均为0 |

1. **特征值定义**



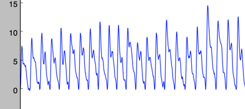
**图4.1 脉搏细节**

1. 峰值谷值（幅值）

峰值是脉搏波形一个周期内的最大值，即上图中主波峰值，谷值是周期内的最小值，通常经过归一化后谷值为0。脉搏幅值=峰值-谷值

1. 周期

周期指脉搏一次有规律的跳动所用时间，定义为两个波谷相隔的时间差。如下图红色箭头所示即一个周期。



**图4.2 脉搏周期**

1. 脉宽

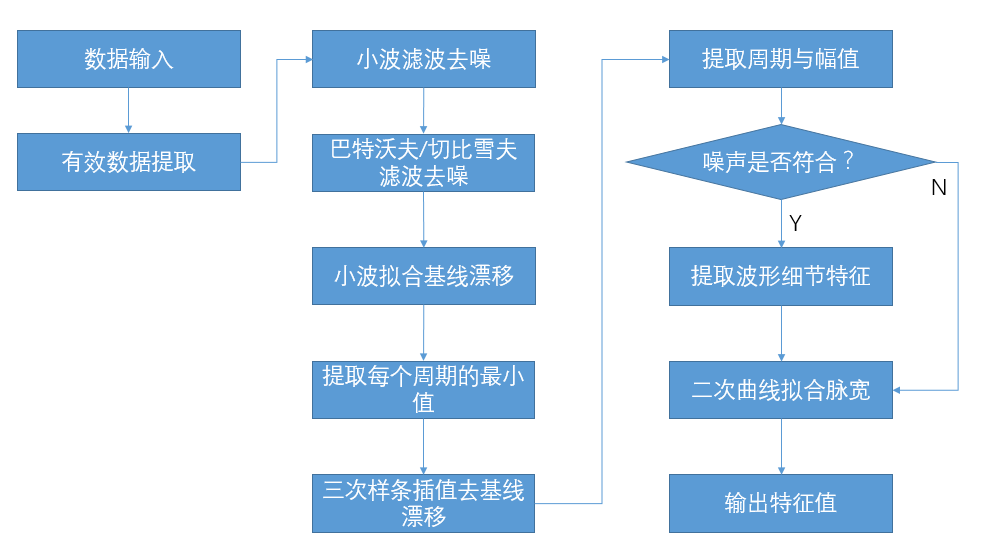
脉宽指手指能够感觉到的脉动的径向范围大小，分为大脉和细脉。定义为一行上六个采样点拟合的二次曲线与x轴的两个交点的距离。

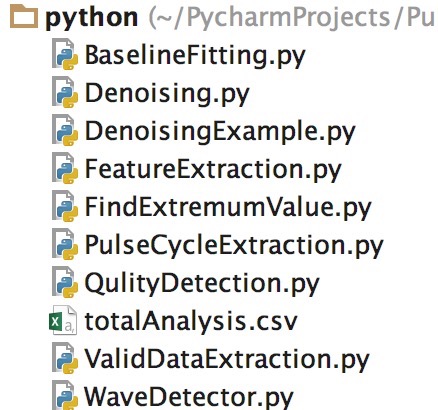
（4） 波形细节特征

在一个周期之内，一个典型的脉搏信号由三个谐波叠加而成，它们分别是主波、潮波和重搏波。主波与潮波之的波谷称为主波峡，潮波与重搏波之间的波谷称为降中峡。这4个特征值构成了脉搏细部特征，如图4.1所示。

1. **程序架构说明**

程序主要分为三部分：有效数据段提取、去噪、特征提取。其架构图如下：





**有效数据提取程序**

接口：ValidDataExtraction根据各个原始数据对应的有效数据起点终点记录数据，将原始数据中对应的有效段按照一定规则筛选。

输入参数：原始数据，有效记录数据，筛选规则

输出参数：有效数据段文件。每个病例数据有寸关尺5行采样点，一行中有0~N段有效的数据段，该程序提取每一个有效数据段作为一个数据文件。

注：有效数据提取与后面的处理分为两个独立工程，目的是为了可以分开运行。

**整体引擎程序**

接口：WaveDetector 整个去噪与特征提取算法的控制逻辑，同时也包括脉宽拟合算法。

输入参数：数据文件，文件格式见有效数据提取的输出

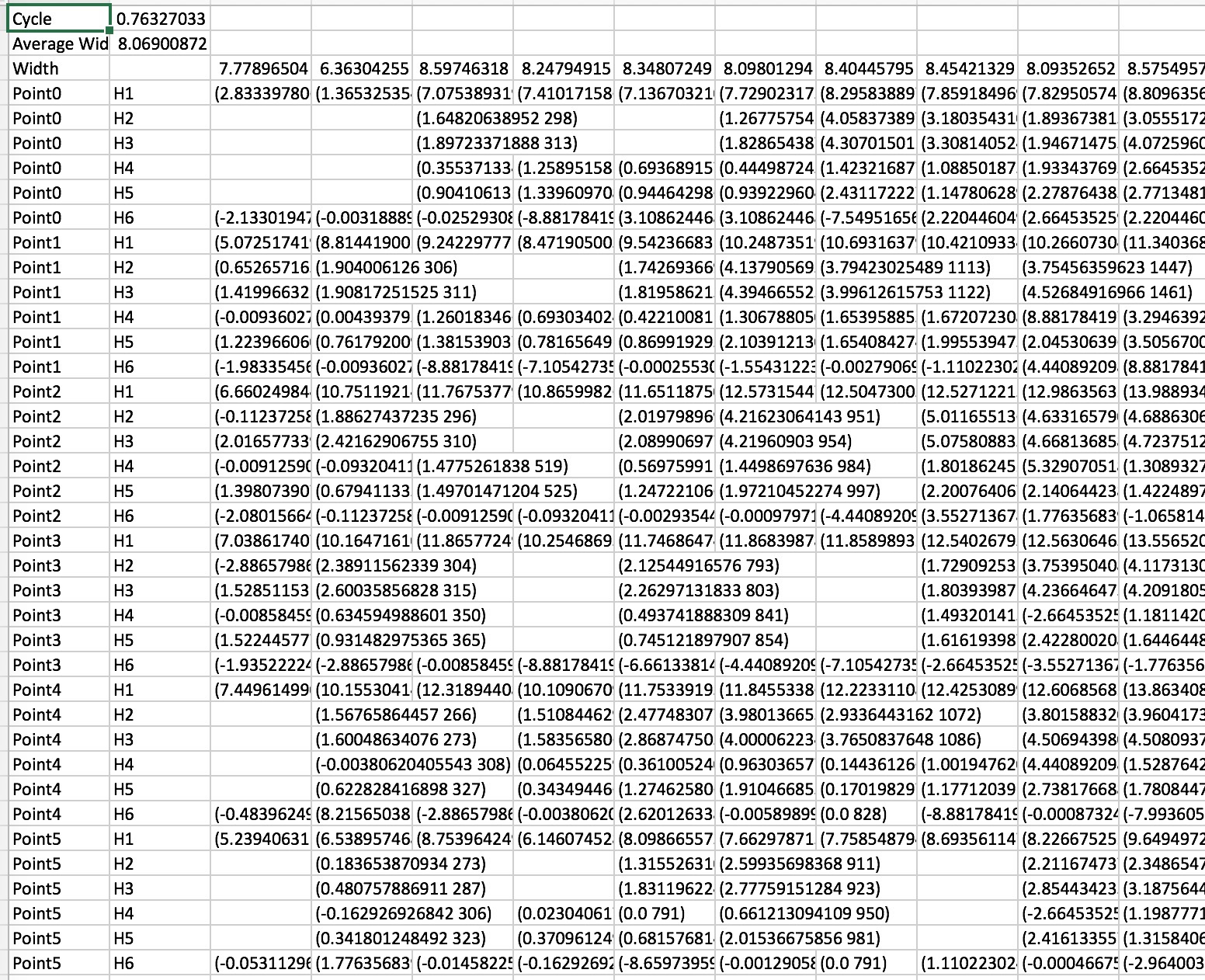
输出参数：提取完特征值的文件，其格式如下：

第一行为周期值，

第二行为平均脉宽，

第三行为每个周期下的波形脉宽，

第四行开始Point-N表示第N个采样点，其后的H1-H6是6个特征值，一次表示为主波幅值、潮波幅值、重搏波幅值、主波峡幅值、降中峡幅值、波谷幅值。每个值由（幅值 出现时间点）组成。



**去噪程序**

接口：Denoising 使用小波进行第一次去噪与去基线漂移，以及使用巴特沃夫与切比雪夫滤波器去噪。

输入参数：一个采样点的原始波形数据

输出参数：去噪后波形数据

接口：BaselineFitting 使用三次样条插值再次去除基线漂移，使得所有波谷点的值都为0。

输入参数：去噪后波形数据

输出参数：去除基线漂移后波形数据

接口：FindExtremumValue 寻找所有周期的最大值与最小值

输入参数：一个采样点的波形数据

输出参数：所有周期的最大值列表与最小值列表

接口：QulityDetection 判断该波形数据是否达到去噪要求

输入参数：波形数据

输出参数：是否达到要求

**特征提取程序**

接口：PulseCycleExtraction 提取波形数据的周期值

输入参数：去噪完的波形数据

输出参数：每个周期的时间长度

接口：FeatureExtraction 提取波形数据的细节特征值

输入参数：去噪完的波形数据

输出参数：所有周期的细节特征值列表，包括主波幅值、潮波幅值、重搏波幅值、主波峡幅值、降中峡幅值、波谷幅值。

1. **算法说明**
2. **有效数据段提取**

由于原始数据中并不是所有的数据都是好的，筛选出符合数据去噪的前提条件的数据可以有效的提高数据的去噪效果。

该过程的输入参数有三个，分别是原始数据，有效数据起点终点记录数据筛选规则。算法可分为三部分：获取原始数据，提取所有有效数据，根据筛选规则进一步提取并把一每段有效数据单独保存。关键在后两个部分。

**提取所有有效数据：**有效数据的起点终点记录文件记录了每一个数据文件每一列的若干个有效区间，区间的含义为从若干行之间为有效段。算法读取区间数据以及对应的原始文件与列，然后提取出对应的原始数据，获得有效数据。

**提取符合筛选规则数据：**上一部分算法已经提取出所有有效数据，但是并不是所有的有效数据都适用于之后的去噪等过程，所以需要进一步制定筛选规则。算法规定有效数据至少有3个点并且长度均大于3000时为最终的有效数据。把满足筛选规则的数据单独作为一个数据文件输出，以便下一个阶段处理。

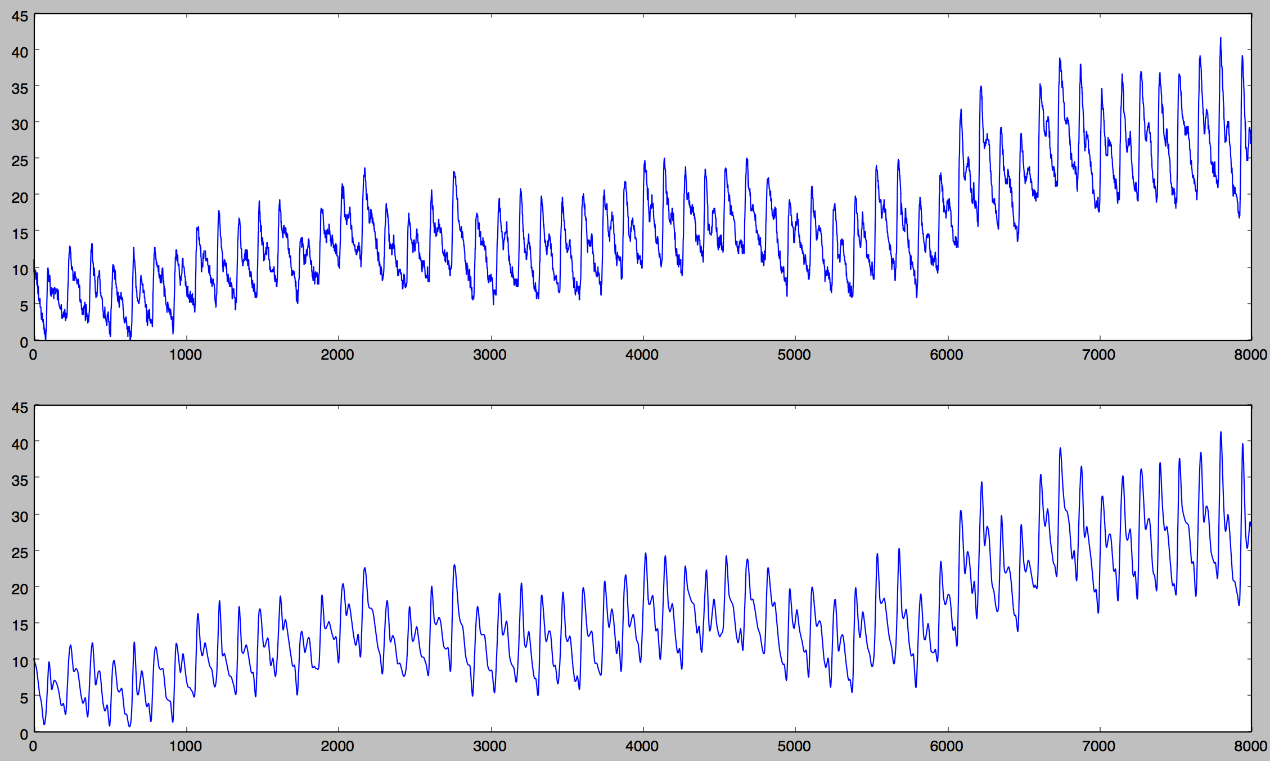
1. **去噪**

脉博波噪声主要包括采集工具本身的噪声，采集过程中因采集外压抖动与人体抖动引入的噪声两部分。此外由于采集对象的呼吸与晃动，以及采集者施压的不均匀，会使得脉博信号产生基线漂移，从而导致不同程度的变形。

由于硬件条件问题，目前采集到的波形噪声较大，因此本文在小波滤波的基础上引入了巴特沃夫带通滤波器以及切比雪夫二型带阻滤波器做进一步的去噪。同时使用三次样条插值拟合基线漂移。

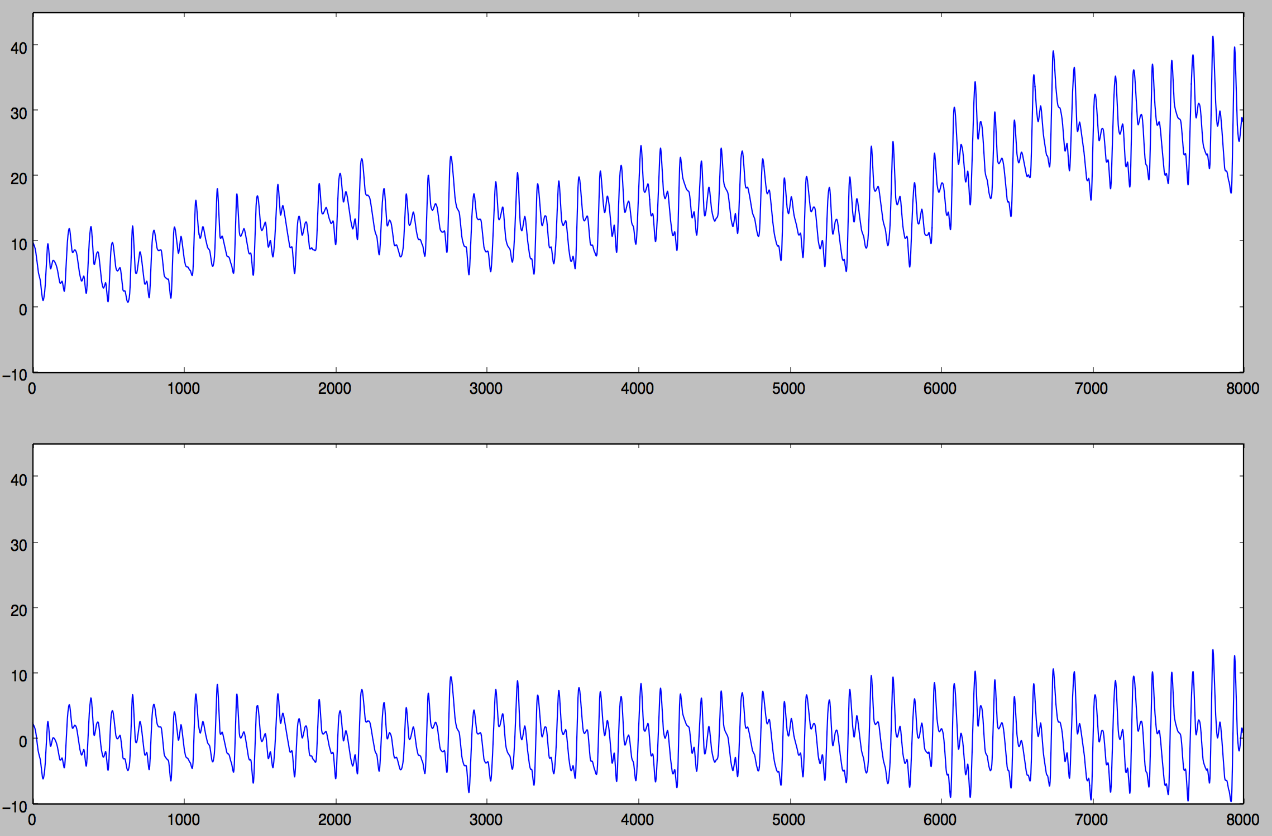
下面将详细讲述这三个滤波器的作用与算法，以及基线漂移去处算法。

**双重小波滤波：**小波变换相比传统傅立叶变换具有正交性，方向选择性，可变的时频域分辨率等特点，适用于脉博信号去噪。不同于前人工作，本文使用了双重小波滤波法进行第一步的去噪与去基线漂移。其中小波基都采用**Sym8**小波，分解层次均为8层。第一次小波分解抛弃1-4层分量，保留5-8层分量进行重组，获得去噪后的波形。下图是去噪效果对比：



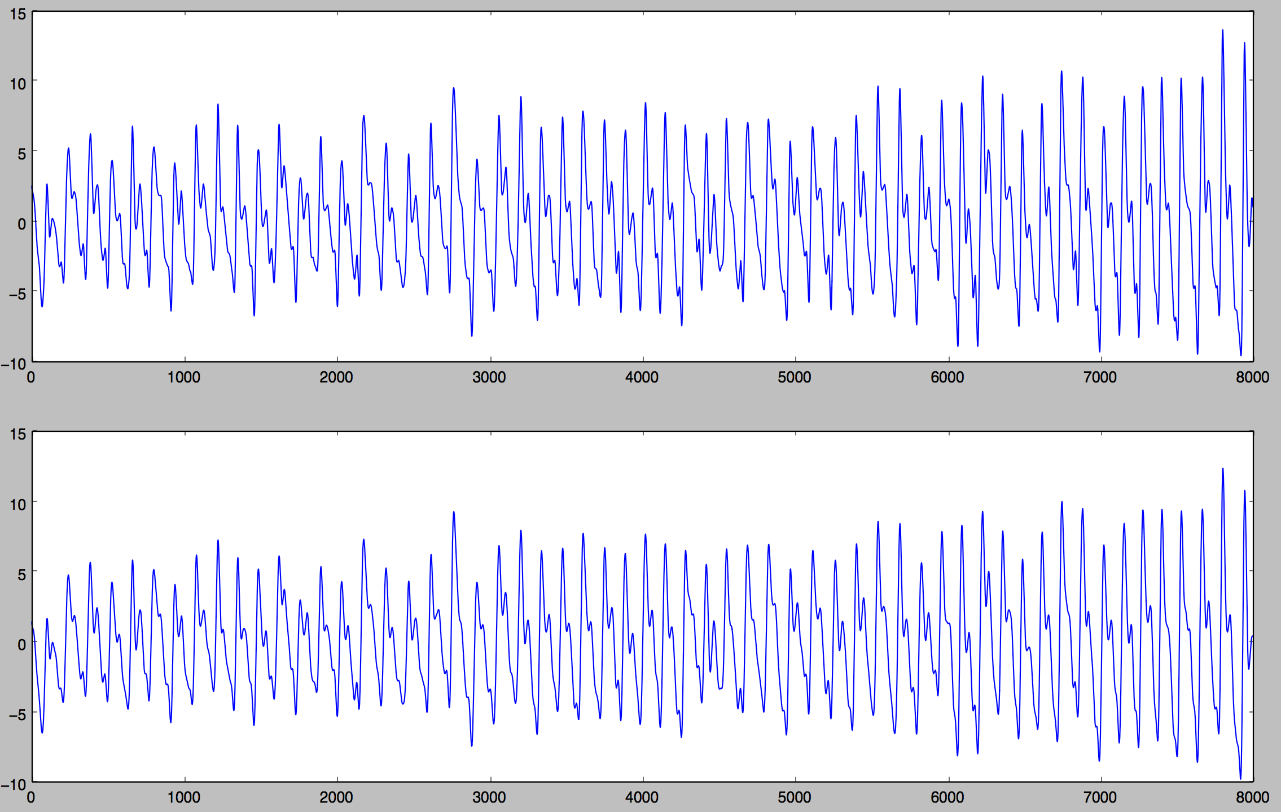
**图4.3 小波滤波效果**

第二次小波分解，保留第0层分量进行重组后作为拟合基线，分解前的波形减去拟合基线即初步去处基线漂移后的波形，如下图所示：



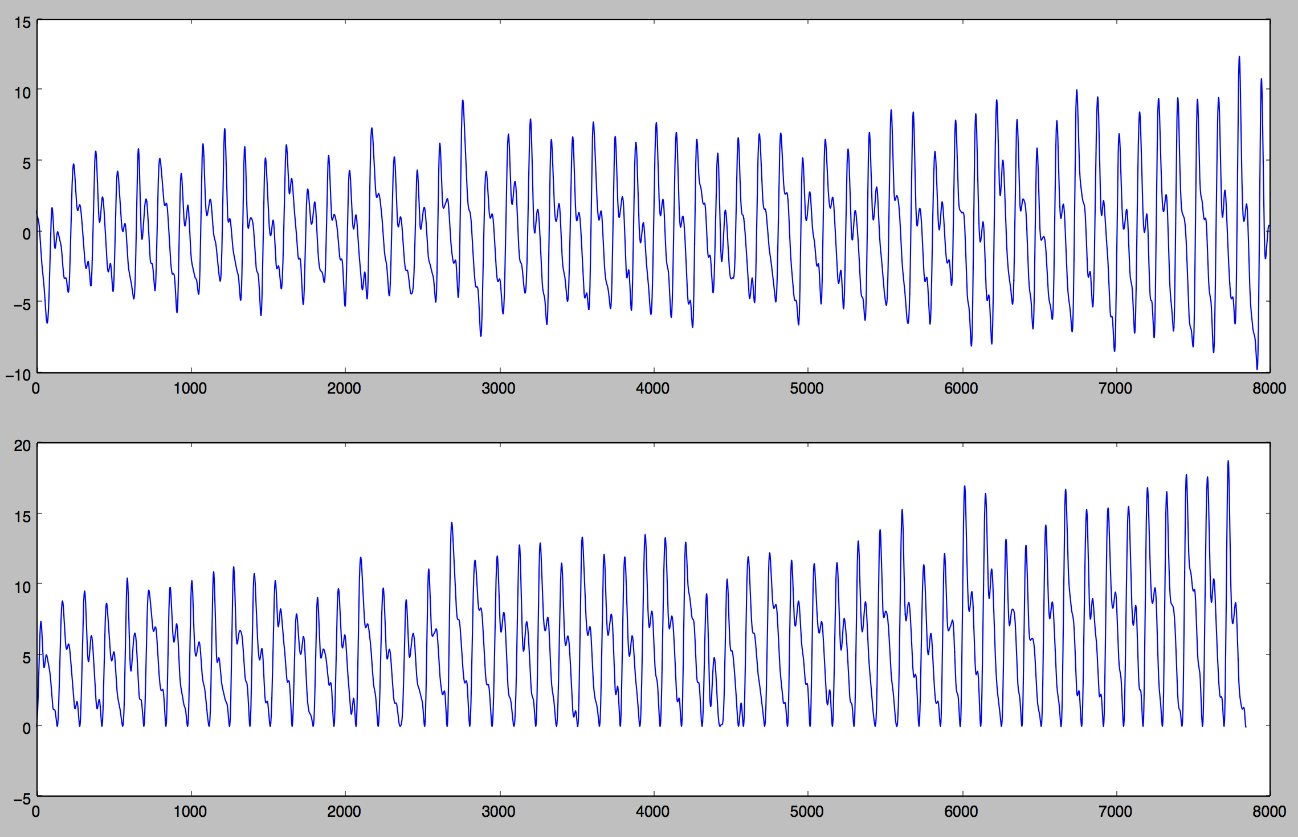
**图4.4 小波去除基线漂移效果**

巴特沃夫带通滤波器**&**切比雪夫二型带阻滤波器：由于脉博波形频率主要分布在5Hz左右，而第一次小波分解无法去除一部分接近脉博波频率的噪声，因此通过巴特沃夫带通滤波器与切比雪夫二型带阻滤波器进行进一步去噪，得到更好的波形。去噪效果如下：



**图4.5 滤波器去噪效果**

三次样条插值去除基线漂移：三次样条插值用于进一步拟合基线值，将每个周期的最小值拟合到0点。效果如下：



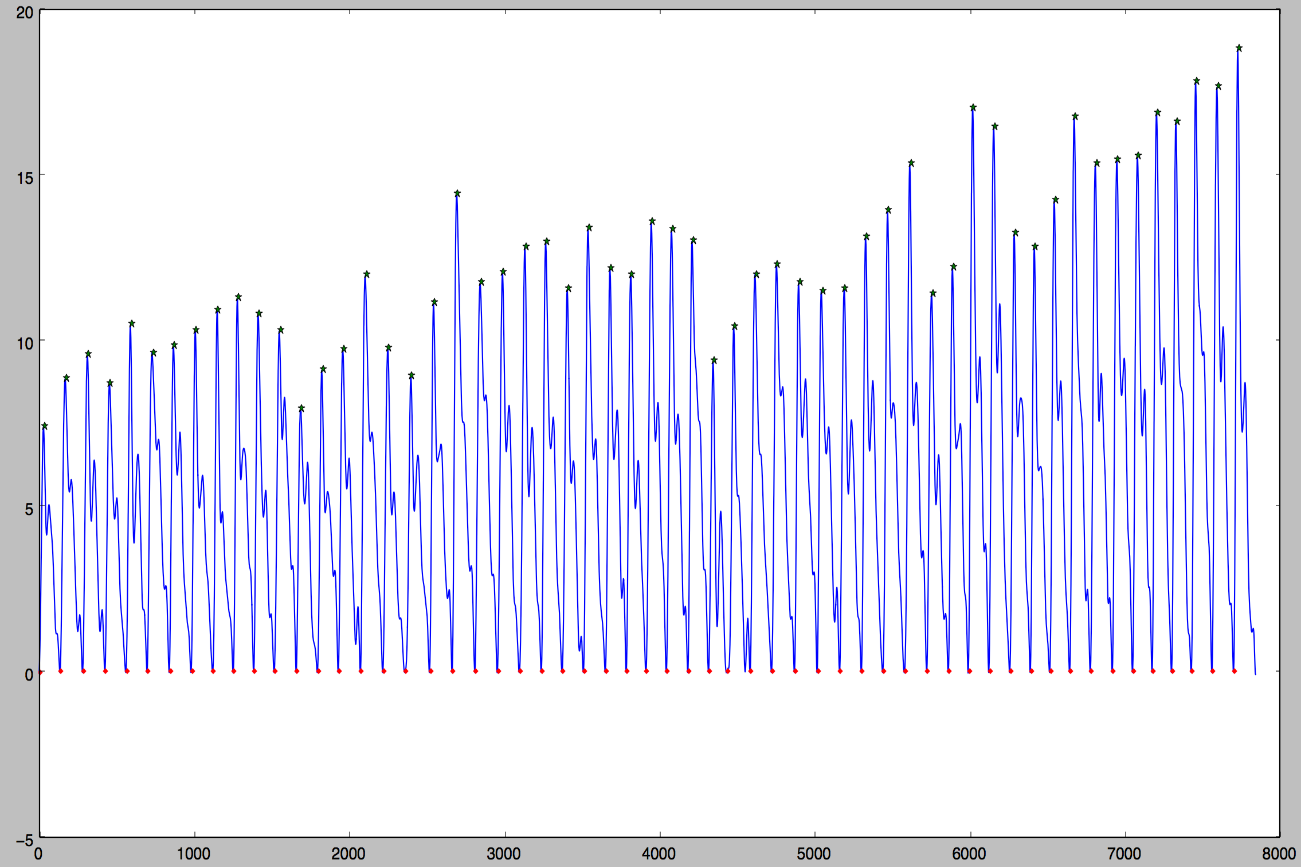
**图4.6 三次样条插值拟合效果**

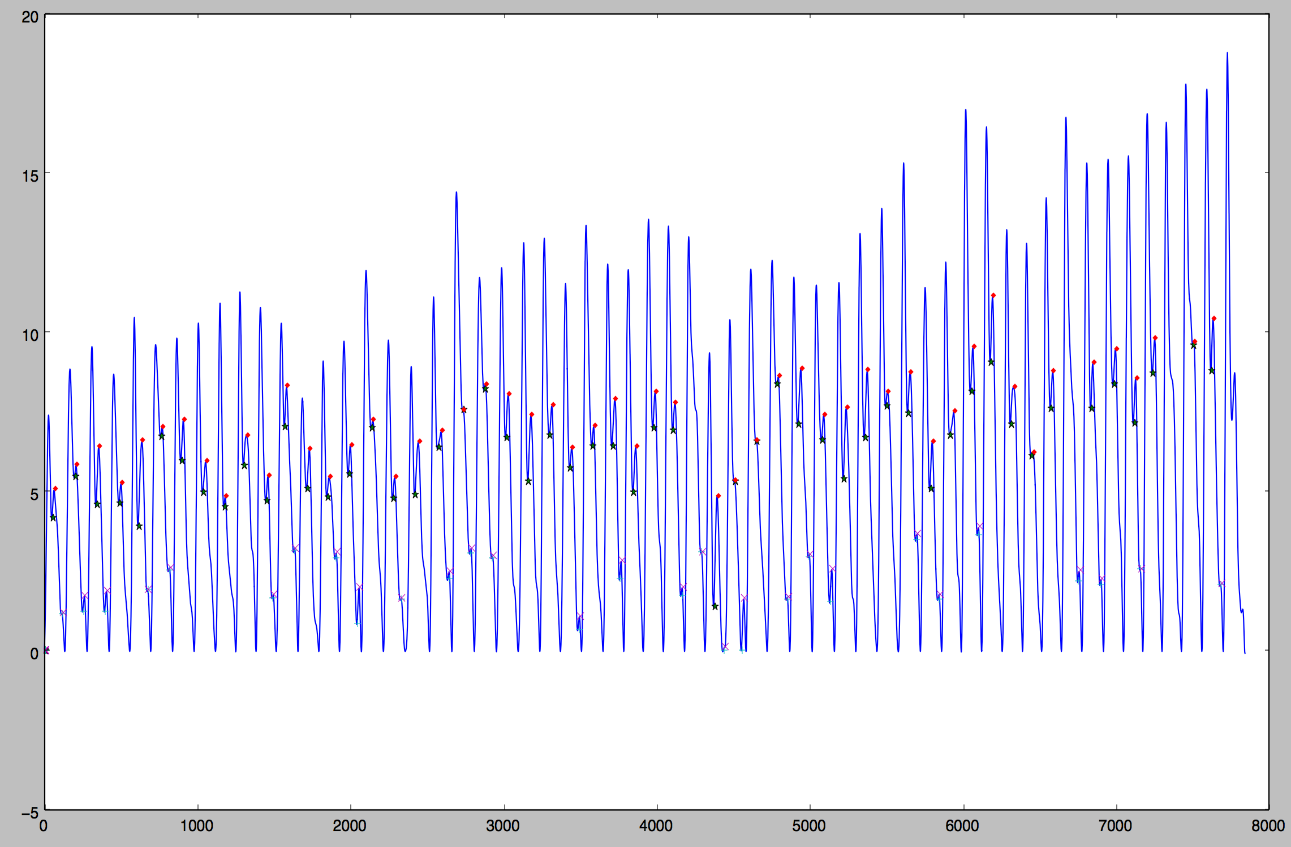
1. **特征提取方法**

特征提取主要分为四部分，分别是：周期/频率，幅值，脉宽，波形细节。

1. **前置算法-局部极值列表提取：**局部极值通过对波形曲线求二阶导数获得，局部极值中包含了所有主波峰点，重搏波峰点等所有波形细节点，但同时也含有一些噪声点。
2. **幅值：**由于人体脉搏波的周期大约在60-120次/分钟，而本文的采样频率为200Hz，因此周期大约在100-200个点之间，这里取100个点为一次估计周期，找到这100个点中的极大值，再与局部极值列表中的点比对，查看是否大于局部极值列表中的左右两点，如果大于，说明是波峰，即得到了该周期的幅值，否则不是。波谷同理。
3. **周期频率：**两个波谷间的点数＊采样频率即该波的周期，对所有波形的周期取平均，可得到整体的周期与脉博频率。
4. **脉宽：**对同一行上4～6个点同一时间的幅值做二次曲线拟合，拟合曲线如果与x轴相交，则两交点的焦距可认为是脉宽，同理对所有周期上的波形采用如上算法并取平均，得到平均脉宽。
5. **波形细节提取：**如上文所说，所有波形细节点都处于局部极值列表中，通常正常脉博波一个周期内拥有2～3个波峰，在时域上3个峰值的间隔时间相对固定，且幅值相差较大。因此本文采用如下算法提取：
   * 1. 找到3个峰值的估计时间间隔范围
     2. 从主波峰开始，如果主波峰后的第二个点处于重搏前波估计时间间隔范围内，说明可能是重搏前波峰，否则是噪声点，继续往后找到所有处于该时间范围内的点。如果找到的点数目大于等于3个，说明噪声过大，放弃该周期的波。
     3. 同理继续往后找到重搏波峰值。
     4. 对于波谷则从主波波谷从后往前找。

由于不需要所有周期的波的波形细节，因此对于噪声相对较大的波形直接忽略，可以获得更好的波形细节特征。下两图分别给出了提取到的波峰波谷值，以及重搏前波与重搏波的特征值：





1. **数据测试**
2. **原始数据情况分析**

原始数据为经过太一有效数据提取程序提取后的数据。总共有140份病例数据，对应4200个采样点。其中存在有效数据的采样点数目为2677个。

详细情况见Excel文件“测试数据情况原始.xls”

1. **有效数据段提取情况分析**

经过有效数据段提取以及按寸关尺拆分后总共有660个有效数据文件。

总体情况如下：

3个点的数量为264；

4个点的数量为215；

5个点的数量为137；

6个点的数量为44。

有效数据段提取详细情况见Excel文件“测试数据情况分析.xls”

1. **去噪与特征提取结果情况分析**

对上述660个有效数据段文件进行特征提取后，有119个文件因噪声原因无法提取特征值，剩余541个可以都可以提取特征值。

针对周期：所有数据都能准确提取周期

针对脉宽：有119个文件因噪声原因无法提取，有183个文件因采样点不够或者采样点幅值存在问题而无法拟合有效脉宽，剩余358个文件可以拟合有效脉宽。

358个脉宽数据中，最小值为2.564，最大值为384.66，分布在［2，16］范围内数据有327个，大于100的数据有5个（认为不合理），去掉5个不合理数据后所有脉宽分布如下图所示：

去噪与特征提取详细情况见压缩包文件“最终运行结果.zip”。其文件格式如下：

**文件名格式：**

*病人原始数据文件名\_寸关尺索引\_有效段索引.csv*

寸关尺索引有1-5，分别代表寸第一行6点，寸第二行6点，关第一行6点，关第二行6点，尺6点。有效段索引为每行6点中第几个有效段，从1开始计数

**每个文件格式：**

第一行 脉宽，单位采样点间距

第二行 周期，单位秒

第三行开始，每6行一个采样点，每个采样点有6行数据，每行数据有所有周期的波形细节及幅值，H1,H2,H3,H4,H5,H6，分别代表：

H1 主波峰幅值，在当前周期内所处位置

H2 主波峡幅值，在当前周期内所处位置

H3 潮波幅值，在当前周期内所处位置

H4 降中峡幅值，在当前周期内所处位置

H5 重搏波幅值，在当前周期内所处位置

H6 波谷幅值，在当前周期内所处位置

1. **典型案例分析与错误数据分析**

典型案例：见Word文件“去噪与特征处理典型案例.doc”

脉宽问题说明：见Word文件“脉宽问题说明.doc”