

# Automated sleep scoring and sleep apnea detection in children

David P. Baraglia<sup>a</sup>, Matthew J. Berryman<sup>a</sup>, Scott W. Coussens<sup>b</sup>, Yvonne Pamula<sup>c</sup>, Declan Kennedy<sup>c</sup>,  
A. James Martin<sup>c</sup> and Derek Abbott<sup>a</sup>

<sup>a</sup>Centre for Biomedical Engineering and School of Electrical and Electronic Engineering,  
The University of Adelaide, SA 5005, Australia.

<sup>b</sup>Department of Physiology, School of Molecular and Biomedical Sciences,  
The University of Adelaide, SA 5005, Australia.

<sup>c</sup>Department of Pulmonary Medicine, Women's and Children's Hospital, 72 King William Road,  
North Adelaide, SA 5006, Australia.

## ABSTRACT

This paper investigates the automated detection of a patient's breathing rate and heart rate from their skin conductivity as well as sleep stage scoring and breathing event detection from their EEG. The software developed for these tasks is tested on data sets obtained from the sleep disorders unit at the Adelaide Women's and Children's Hospital. The sleep scoring and breathing event detection tasks used neural networks to achieve signal classification. The Fourier transform and the Higuchi fractal dimension were used to extract features for input to the neural network. The filtered skin conductivity appeared visually to bear a similarity to the breathing and heart rate signal, but a more detailed evaluation showed the relation was not consistent. Sleep stage classification was achieved with an accuracy of around 65% with some stages being accurately scored and others poorly scored. The two breathing events hypopnea and apnea were scored with varying degrees of accuracy with the highest scores being around 75% and 30%.

**Keywords:** Sleep, apnea, EEG, neural network

## 1. INTRODUCTION

### 1.1 Objectives

This paper investigates the development of software capable of automatically detecting sleeping disorders in children. Three tasks have been proposed for the software:

1. the measurement of depth of sleep of a patient,
2. the detection of breathing irregularities during sleep, and
3. the extraction of heart and breathing signals of a patient.

The first task was to investigate the use of filtering techniques on the skin conductivity to extract the heart rate and breathing rate from only the skin conductivity. The second task was to develop software capable of accurately scoring a patient's sleep. This task investigated the effectiveness of the two signal processing techniques of Fourier transform and fractal dimension in the classification of sleep stages. The third task investigated the detection of breathing events using only the EEG as input data. The same two signal processing techniques as in the second task were used in this task as well.

Data sets have been provided from the Sleep Disorders Unit at the Adelaide Women's and Children's Hospital. The data includes a set of electroencephalograph (EEG), thoracic and abdominal breathing data and skin conductivity data. By using various signal processing techniques on the data it is hoped that the depth of sleep of a patient can be accurately estimated as well as the detection of occurrences of events during sleep where breathing is obstructed. It is hoped that the results of this project will contribute towards the development of automated systems for the diagnosis of sleep apnea in children.

## 1.2 Motivation

Overnight polysomnography (PSG) in a sleep laboratory is considered to be the definitive diagnostic exam for suspected abnormal breathing in children<sup>1</sup>. Polysomnography is the simultaneous recording of a number of physiological variables of a patient over a nights sleep. Typically they include measurements of brain activity, respiratory activity, oxygen levels, cardiac activity and muscle activity.

One of the most common sleep disorders is the Obstructive Sleep Apnea Syndrome (OSAS) with a prevalence of 2% in children aged two to 18 years<sup>2</sup>. OSAS is caused by the repeated cessation of airflow due to upper airway obstruction. OSAS is usually defined as more than five breathing events per hour of sleep<sup>3</sup>. A breathing event during sleep is an event in which the patients breathing stops or is reduced. They are classified as being either an apnea or a hypopnea. An episode of apnea may be defined as the temporary absence of breathing for at least two respiratory cycles while hypopneas may be defined as a reduction of more than 50% in respiratory movement for at least two respiratory cycles<sup>4</sup>.

While the PSG is considered the diagnostic standard for the diagnosis of OSAS there are considerable drawbacks to their use. The PSG is uncomfortable for the patient and involves a considerable investment for the healthcare system requiring equipment, bed space and specialized technical support<sup>4</sup>. Interpretation of data is a time consuming and tedious process requiring a qualified sleep technician. Thus the automation of tasks involved in sleep studies is a highly desirable goal. The advantages of an automated system for diagnosis include speed, reliability economic savings and improved reliability of diagnosis<sup>5</sup>. If important information can be automatically obtained from an EEG it would be possible to develop portable devices that could be taken to a child's home. Compared to a full sleep study this would be much less stressful on the child and would be less expensive. Portable home sleep monitoring has been proposed for the diagnosis of OSAS. The potential benefits of home sleep monitoring include reduced costs and the ability to evaluate patients in their usual environment<sup>6</sup>.

## 2. SKIN CONDUCTIVITY

### 2.1 Data

The data provided for this study was recorded over the duration of a night for four patients. The data consists of the galvanic skin conductivity together with the corresponding thoracic breathing signal and electrocardiogram. The skin conductivity was sampled at 100 Hz, the thoracic data at 25 Hz and the ECG at 500 Hz.

### 2.2 Procedure

#### 2.2.1 Overview

The skin conductivity is processed by applying a bandpass filter centred on the frequency of interest, which was either the breathing rate or the heart rate. Following this a visual comparison of the filtered skin conductivity with either the breathing or heart signals is possible, but in order to quantify any resemblance that may appear between the signals their correlation is calculated. When the patient is breathing regularly their thoracic signal has a sinusoidal appearance and so does not require any processing but the ECG has a completely different shape. An algorithm is used to convert the ECG into a square wave signal representing the heart rate.

#### 2.2.2 Heart rate detection algorithm

The QRS detection algorithm—designed specifically for the production of heart rate from an ECG—was used. The term QRS refers to points in the general shape of the ECG. This shape consists of a relatively fast, high amplitude pulse followed by a slower, lower amplitude pulse. A linear phase bandpass filter with pass band of 10-20 Hz is used to remove the slower pulse. The algorithm then uses three concurrent moving windows of lengths 380 ms, 280 ms and 100ms over the filtered signal. When the large pulse lies in the central window, the other two windows will have little activity in them. The ratio of the activity in the central window to the activity in the other two windows is used to detect heart beats. When this ratio exceeds a certain threshold the output is held at '1' for 280 ms to indicate a pulse, otherwise the output is set to '0'. The operation of the QRS detection algorithm is shown in Figure 1. The first and second graphs show the ECG input signal and the filtered ECG. The third graph shows the pulse factor the ratio used in detecting

pulses. The dashed line in this graph represents the threshold used for pulse detection. The last graph shows the output of the QRS algorithm, representing the heart rate of the ECG.

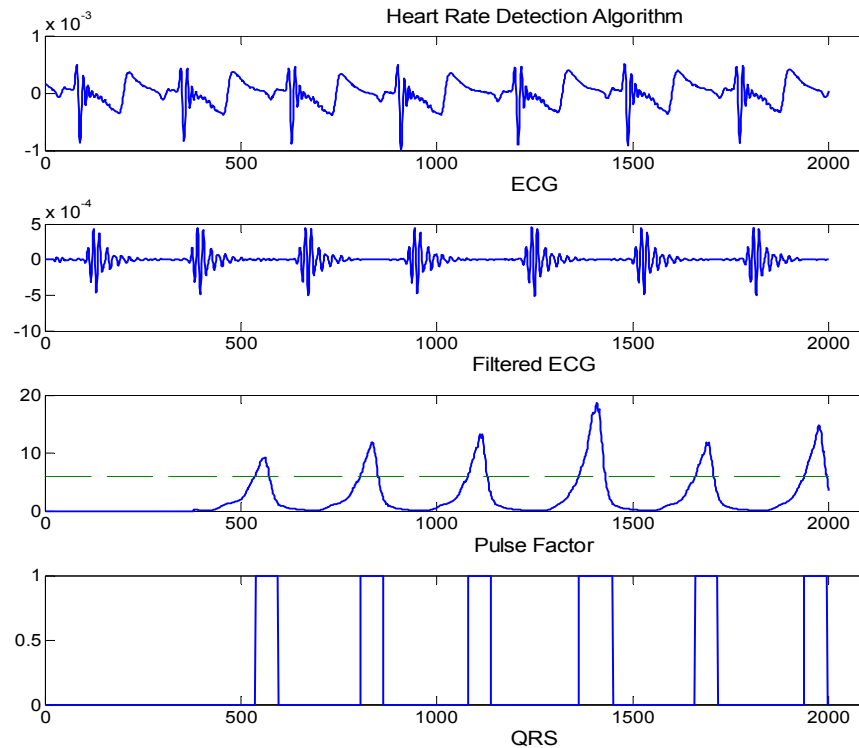


Figure 2.1: This figure shows the input to and outputs of the stages of the QRS detection algorithm. The top graph is the raw, unprocessed ECG signal. We first filter the ECG with a linear phase bandpass filter, and the output is shown below the ECG. Then applying three sliding windows to the filtered ECG signal gives the pulse factor. When the pulse factor moves above a threshold (shown as the horizontal line) a pulse is registered in the QRS signal. The time delays between the original pulse and the final pulse are unimportant as we then compensate for this in the correlation method we use.

### 2.2.3 Skin conductivity filtering

Bandpass filtering was used to extract the frequency content of the skin conductivity at the frequencies of interest. The breathing rate of children is around 20-30 breaths per minute for preschool children and around 16-25 breaths per minute for older children<sup>7</sup>. These correspond to a range of frequencies from 0.267–0.5 Hz. The heart rate for children aged 1-10 years is around 70-120 beats per minute and for children age over 10 up to adults the heart rate is around 60-100 beats per minute<sup>8</sup>. These correspond to the frequency range 1-2 Hz.

Two bandpass filters were tested for both the breathing and heart rate experiments. The two filters used were a Butterworth filter, chosen as it is maximally flat in the passband and a Chebyshev type II filter, chosen as it is monotonic in the passband. Both filters were fourth order. For the breathing experiment a passband of 0.3-0.5 Hz was found to be effective while for the heart rate experiment the frequency range 1-2 Hz was used as the passband.

### 2.2.4 Correlation

When correlating the skin conductivity with either of the two other signals one must take into consideration that there may be a delay between the two signals. Therefore when calculating the correlation we first apply a time shift to one of the signals. For a signal  $x$  and a time shift of  $m$  samples let  $x_m(n)=x(m+n)$  define the time shifted signal  $x_m$ . The correlation of signals  $x$  and  $y$  is now a function of the time shift given by

$$r(m) = \frac{\text{cov}(x_m, y)}{\sqrt{\text{cov}(x_m, x_m) \text{cov}(y, y)}}, \quad (1)$$

where  $\text{cov}(a, b)$  is the covariance of discrete time sequences  $a$  and  $b$ . By means of calculating the correlation for a range of time shifts, a time shift that maximizes the correlation is sought.

## 2.3 Results

### 2.3.1 Breathing rate

The correlation function was calculated for each of the four patients and for each of the two filters at seven times during the night. In each case the correlation was calculated for the duration of 30 seconds. Figure 2.2 shows an example of the skin conductivity before and after filtering along with the breathing signal.

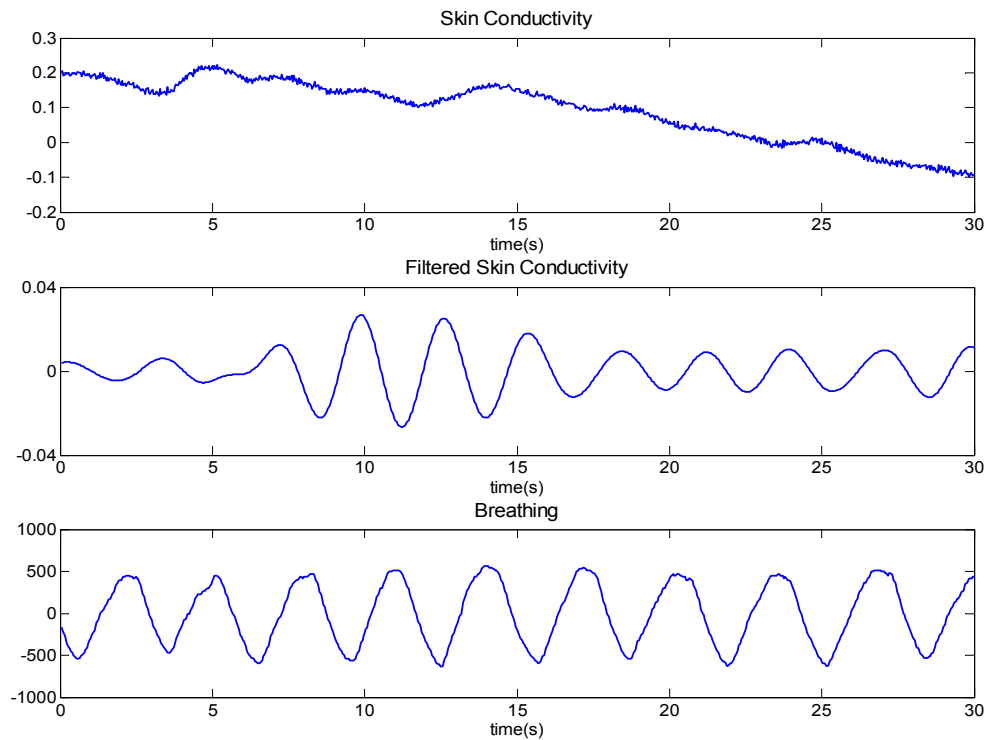


Figure 2.2: The original unprocessed skin conductivity signal, filtered skin conductivity and breathing signals.

The correlation function for the filtered skin conductivity and breathing of Figure 2.2 is shown in Figure 2.3. The minimum and maximum values of the correlation and the corresponding time shifts are shown in Table 2.1 in the form of  $\text{mean} \pm \text{one standard deviation}$ . The peak values of correlation vary notably between patients, being highest for patient 1. This suggests that the correlation varies from person to person. Examining the breathing signals of the four patients showed that patient 1 had the most regular breathing while patients 2 and 3 showed irregular breathing due to their breathing disorders. Patients 2 and 3 also happen to have the lowest correlation. It is reasonable to expect that patients with regular breathing are going to have the highest correlation because the bandpass filter used was centred around the frequency for regular breathing.

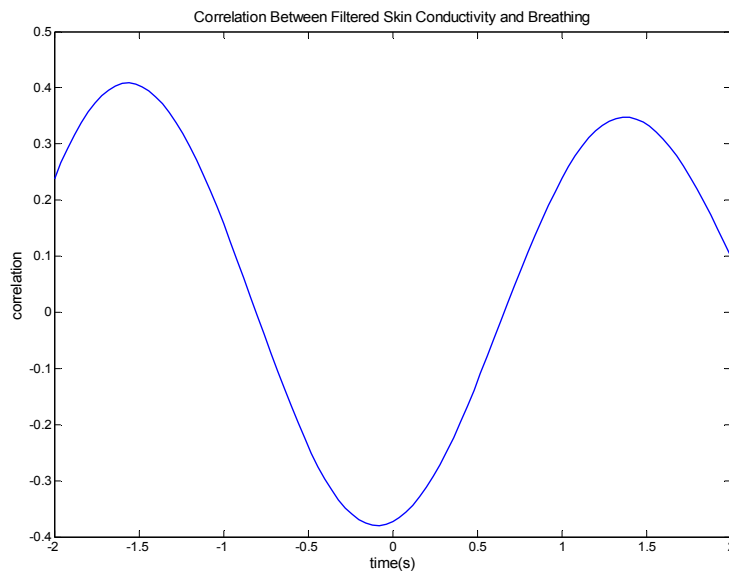


Figure 2.3: Correlation between filtered skin conductivity and breathing.

Although the larger values of correlation obtained are large enough to suggest some relation between the signals, there is substantial variation between the positions of the correlation peaks from one time to another. This complete lack of consistency suggests that the visual resemblance between the filtered skin conductivity and breathing signals is due only the application of a filter centred around the breathing rate and not due to an influence of breathing on skin conductivity.

Table 2.1: Correlation statistics for skin conductivity and breathing.

Filter 1: Butterworth

Patient	1	2	3	4
Maximum correlation	$0.40 \pm 0.17$	$0.12 \pm 0.07$	$0.20 \pm 0.13$	$0.24 \pm 0.14$
Time shift (s)	$-0.11 \pm 0.61$	$-0.06 \pm 1.04$	$-0.13 \pm 1.22$	$0.29 \pm 0.90$
Minimum correlation	$-0.32 \pm 0.17$	$-0.14 \pm 0.10$	$-0.24 \pm 0.9$	$-0.24 \pm 0.15$
Time shift (s)	$0.01 \pm 1.00$	$0.18 \pm 1.22$	$0.03 \pm 0.97$	$0.09 \pm 1.23$

Filter 2: Chebyshev type II

Patient	1	2	3	4
Maximum correlation	$0.51 \pm 0.23$	$0.17 \pm 0.06$	$0.17 \pm 0.04$	$0.30 \pm 0.13$
Time shift (s)	$0.70 \pm 0.94$	$0.2 \pm 0.93$	$0.77 \pm 0.81$	$0.24 \pm 0.98$
Minimum correlation	$-0.48 \pm 0.19$	$-0.17 \pm 0.08$	$-0.14 \pm 0.15$	$-0.28 \pm 0.12$
Time shift (s)	$0.00 \pm 0.93$	$0.26 \pm 1.18$	$0.01 \pm 1.07$	$0.08 \pm 1.05$

### 2.3.2 Heart rate

Like the breathing experiment the correlation function was calculated for each of the four patients and for each of the two filters at seven times during the night. It was found that the performance of the Chebyshev II filter was notably poorer than the Butterworth so only the results for the Butterworth filter as presented. The correlation was calculated for the duration of 15 seconds. Figure 2.4 shows an example of the skin conductivity before and after filtering, along with the heart rate.

The correlation function for the filtered skin conductivity and heart rate of Figure 2.4 is shown in Figure 2.5. The peak values of correlation and the corresponding time shifts are shown in Table 2.2. The peak values of correlation are generally quite low and once again show variation from person to person. Again there is substantial variation between the position of the correlation peaks from one time to another indicating that there is no clear time delay between the signals.

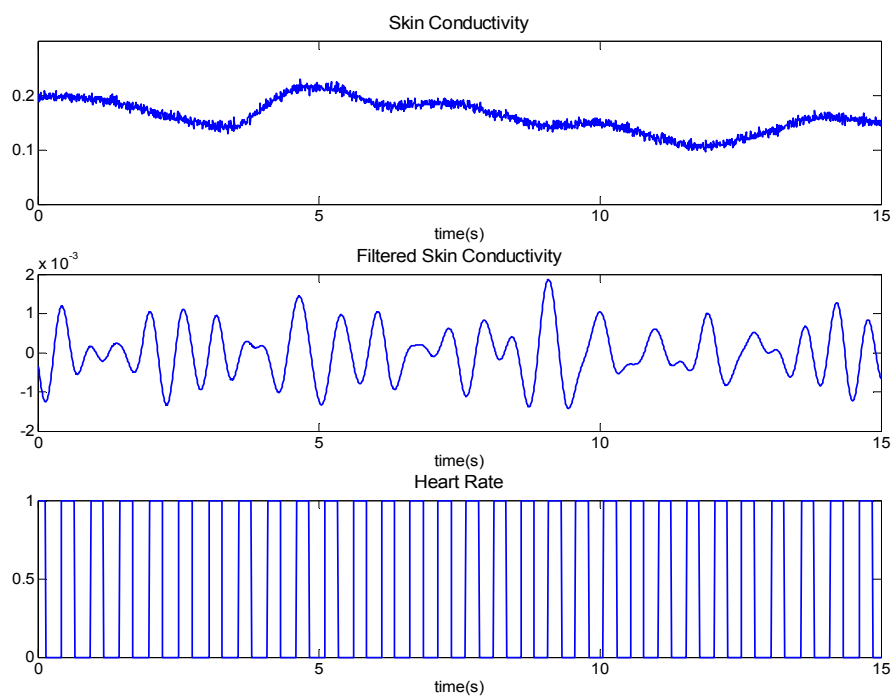


Figure 2.4: The unprocessed skin conductivity signal, the filtered version, and heart rate as extracted from the original ECG signal using the QRS algorithm mentioned in Subsection 2.2.2.

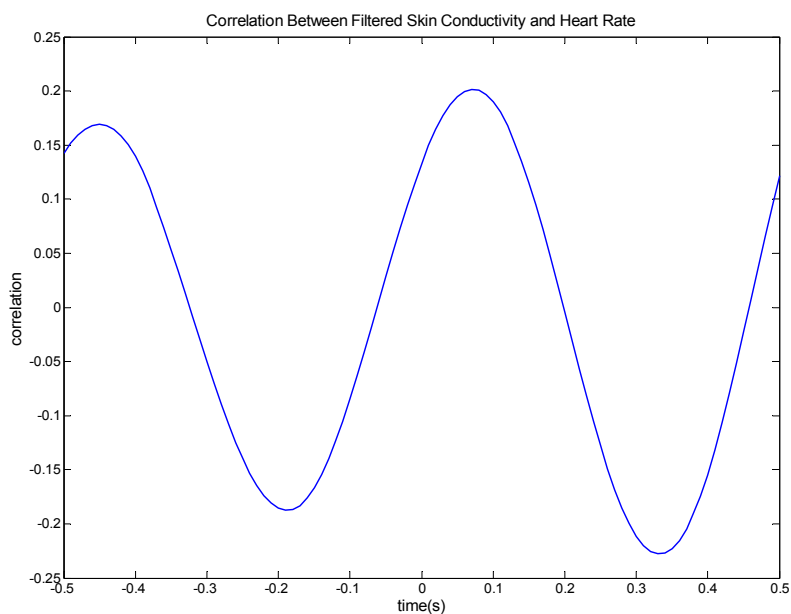


Figure 2.5: Correlation between filtered skin conductivity and heart rate.

Table 2.2: Correlation statistics for skin conductivity and heart rate.

Patient	1	2	3	4
Maximum correlation	0.09±0.06	0.12±0.05	0.24±0.07	0.18±0.10
Time shift (s)	0.01±0.26	-0.06±0.26	0.03±0.27	0.12±0.18
Minimum correlation	-0.09±0.07	-0.11±0.05	-0.25±0.08	-0.18±0.11
Time shift(s)	0.15±0.26	-0.03±0.24	0.15±0.19	-0.02±0.29

### 3. SLEEP STAGE SCORING

#### 3.1 Data

The data consists of a night's worth of EEG data and corresponding hypnograms for 26 patients. Of the 26 patients 13 of them are healthy subjects referred to as the control patients while the other 13 suffer from breathing disorders and are referred to as the active patients. The sampling rate for the EEG data was 125Hz for patients 1-24 and 250Hz for patients 25 and 26. A hypnogram contains human-classified sleep stages based on the PSG. The hypnograms were scored using 30 second epochs and sleep stages were classified into 7 types: awake, stages 1-4, REM and movement.

#### 3.2 Procedure

##### 3.2.1 Overview

A neural network is used to score sleep stages from the input data. The input data for the network is extracted from the EEG signals. Two techniques were used to produce input data: the Fourier transform and the Higuchi estimate of the fractal dimension. Since the sleep stages were scored over 30 second intervals, input data for the network was extracted from the EEG using the same 30 second intervals.

##### 3.2.2 Fourier transform

By applying the Fourier transform to the EEG signal, spectral information of the signal is obtained. The spectrum is then divided into frequency bands and the normalized total power in each frequency band is used as input data for the neural network.

The FFT algorithm is applied to the EEG signal  $x(t)$  over each epoch to produce the Fourier transform  $F(f)$ . The power spectrum is then simply given by  $P(f) = |F(f)|^2$ . As the fast Fourier transform algorithm is fastest for powers of two the number of points used was the next largest power of two after 30 seconds worth of samples. Hence 4096 points were used for the 125 Hz EEGs and 8192 points for the 250 Hz EEGs. The power spectrum is then divided into the following frequency bands; delta 0.5 – 4 Hz, theta 4 – 8 Hz, alpha 8 – 13 Hz, and beta 13 – 30 Hz corresponding to the classification of brain waves. To obtain the total power in each of these frequency bands the power spectrum is summed over the points lying in the corresponding frequency range. The normalized total power in each frequency band is just the power in that frequency band divided by the total power over all frequencies.

##### 3.2.3 Fractal dimension

Fractal dimension is a non-linear measure of a signal's complexity. It is a value between 1 and 2 with values near 1 representing a state of low complexity or linearity while values near 2 represent complex non-linear behaviour<sup>9</sup>. The fractal dimension is calculated over each epoch using the Higuchi fractal algorithm. The algorithm is performed by calculating  $L(k)$ , a measure of the mean length of curve. Let  $N$  be the number of data points. For  $m = 1, 2, \dots, k$  the a measure  $L_m(k)$  of the length of the curve is given by

$$L_m(k) = \sum_{i=1}^{\lfloor (N-m)/k \rfloor} |X(m+ik) - X(m+(i-1)k)|. \quad (2)$$

From this the mean length  $L(k)$  is given as the mean of the  $L_m(k)$  together with a normalization factor so that

$$L(k) = \sum_{m=1}^k \frac{N-1}{\lfloor (N-m)/k \rfloor k^2} L_m(k), \quad (3)$$

for  $k = 1, 2, \dots, k_{\max}$ . If the  $L(k)$  are proportional to  $k^{-D}$  for some  $D$  then the curve is fractal-like with  $D$  the fractal dimension. Thus when  $\log(L(k))$  is plotted against  $\log(k)$  the data should lie on a straight line with slope  $-D$ . Therefore a least-squares linear regression is used to fit a line to the data points  $(\log(k), \log(L(k)))$  and the negative of the slope is taken to be the fractal dimension of the curve. Letting  $x = \log(k)$ ,  $y = \log(L(k))$  then the fractal dimension is given as

$$d = -\frac{\text{COV}(x, y)}{\text{COV}(x, x)}. \quad (4)$$

In implementing the algorithm there remains the choice of  $k_{\max}$ . It was determined by experiment that only a small number of points was needed to produce a good result and so  $k_{\max} = 8$  was used.

### 3.2.4 Network architecture

There are many choices to consider when choosing the network architecture. Once the general type of network has been chosen the main choices to make are the number of layers to use, the number of nodes in each layer and the transfer function to use for each layer. To classify sleep stages seven output nodes were used corresponding to the seven sleep stages to be scored. Each output node produces a value between 0 and 1. Ideally one node should take the value 1 to indicate the sleep stage and the other outputs should be zero. In practical neural networks, the outputs take on a range of 0 to 1. When using the network to score sleep stages the node with the highest output is chosen as the present sleep stage.

A multilayer feedforward network using a back propagation training algorithm was chosen for their effectiveness in classification tasks. When choosing the number of layers to use it should be taken into consideration that at least two layers should be used because two layers are required in order to approximate all well behaved functions arbitrarily well. On the other hand using too many nodes can lead to poor network performance because there is too much freedom. Three layers of neurons were chosen because it was found to be the most effective experimentally. Two choices of node configurations were chosen for the experiment. The choices of nodes for the first layer, second and the output layer were 7, 14 and 7 for the first network and 13, 10 and 7 for the second network. These were found after some experimentation to be suitable node structures. Note that the number of nodes in the input layer is constrained by the number of inputs available, and the number of output nodes is constrained by the number of classification categories.

The transfer function used for the first two layers was the 'tansig' function and for the output layer the 'logsig' function was used. The choice of these functions is due firstly to their output lying between -1 and 1 for 'tansig' and between 0 and 1 for 'logsig' and secondly because these functions are infinitely differentiable they are suitable for backpropagation algorithms.

### 3.2.5 Network Training

The resilient backpropagation algorithm<sup>10</sup> (RPROP) was chosen because it is the fastest of the back propagation algorithms on pattern recognition problems. The training algorithm was to minimise the mean squared error between the desired output and the network output. The training process consisted of 1000 iterations of the algorithm on each training input.

From the EEG and hypnogram data, four sets of training and testing data consisting of the spectral data, the fractal dimension and (in the case of the training data) the human-classified sleep stage were produced. For each of these four sets the data of 13 patients was used as training while the other 13 were used for testing. In the first data set the training data consisted of the active patients and the test data was the control patients. For the second set the control patients were used for training and the active patients for testing. The last two training sets used a mixture of active and control patients for both training and testing. In addition to this, two methods were used to pick out training data from the 13 patients used. The first method was to simply use every fourth epoch as training data, the second method was to pick out roughly equal amounts of each sleep stage in order to produce evenly distributed training data. Finally there is also the choice of using the spectral information alone or using the spectral information and the fractal dimension together.



### 3.3 Results

#### 3.3.1 Overall performance

The experiment was conducted for each choice of network configuration, with and without fractal dimension and with or without using evenly distributed the training data. For each of these choices the network was trained and tested on the four sets of training and testing data (training data was also used to test performance). The mean and standard deviations for the network accuracy is shown in Table 3.1. A number of conclusions can be drawn from Table 3.1. Firstly the accuracy is significantly lower when the training data has been equally distributed. This can be explained by the fact that certain stages, particularly stage 2 and stage 4 are scored with higher accuracy than other stages and these stages also happen to be more common during sleep. Choosing equal amounts of each stage means choosing less of stage 2 and 4 and choosing more of the difficult to score stages—1, 3 and M—hence the overall accuracy is reduced.

While there is surprisingly little difference in performance using the node structures, it is interesting to see that the 13,10,7 node layout consistently achieves a higher accuracy on the training data yet on the testing data, the 7,14,7 node layout consistently achieves a higher accuracy. Finally the addition of the fractal measure appears to only slightly improve the accuracy. One possible explanation for this is that the fractal measure is good for measuring local occurrences of chaotic behaviour in the EEG but over a 30 second epoch it does not provide a useful measure.

Table 3.1: Overall sleep stage scoring accuracy for the neural networks.

Node structure (layer1, layer2, layer3)	Fractal measure used (Y/N)	Evenly distributed training data (Y/N)	Accuracy on training data (%)	Accuracy on test data (%)
7,14,7	Y	N	$71.6 \pm 2.1$	$65.7 \pm 2.0$
7,14,7	N	N	$70.6 \pm 3.1$	$66.0 \pm 3.4$
13,10,7	Y	N	$72.7 \pm 2.7$	$65.2 \pm 2.4$
13,10,7	N	N	$71.1 \pm 3.8$	$65.7 \pm 1.9$
7,14,7	Y	Y	$60.8 \pm 4.3$	$45.2 \pm 2.8$
7,14,7	N	Y	$57.6 \pm 4.4$	$44.4 \pm 2.6$
13,10,7	Y	Y	$61.6 \pm 5.6$	$42.9 \pm 4.5$
13,10,7	N	Y	$59.5 \pm 4.4$	$42.1 \pm 3.4$

#### 3.3.2 Performance of individual training sets

The data for this experiment consisted of a mixture of active and control patients. In this Section we consider the influence this has on the system accuracy. Table 3.2 shows the accuracy of the networks on the four different training and testing data sets. Only the results obtained without evenly distributing the training data was used for these results. Notice that the networks could achieve higher accuracy on the training data when the patients were controls, yet on the testing set the performance is poorer indicating that training with control patients does not prepare the network as well for active patients. Another observation is that the standard deviations are much lower than in Table 3.1 indicating that the performance of the network is more dependent on the training and testing data than on the chosen architecture.

Table 3.2: Network accuracy for each of the training and testing data sets.

Data set	Train with	Test with	Accuracy on training data (%)	Accuracy on test data accuracy (%)
1	Active	Control	$68.1 \pm 1.2$	$67.9 \pm 1.3$
2	Control	Active	$74.8 \pm 0.7$	$62.5 \pm 0.7$
3	Mix	Mix	$72.2 \pm 0.6$	$65.3 \pm 0.6$
4	Mix	Mix	$71.0 \pm 1.0$	$67.0 \pm 0.4$

## 4. BREATHING EVENT DETECTION

### 4.1 Data

The data for this experiment consisted of EEG data from nine patients together with a list of the breathing events for each patient. The list of breathing events specifies for each event the time of occurrence, the duration, the type of event, the sleep stage it occurred in.

### 4.2 Procedure

A neural network is used to perform the automatic detection of breathing events.

As in the previous experiment a multilayer feedforward network using a backpropagation training algorithm was chosen. Although some experimentation was used to obtain node structures with good performance, the issue of node structure was given less consideration than the structure of the input data. While sleep stages are usually scored over 30 second intervals, breathing events can be very brief so the choice of interval duration is an important issue to be investigated. Three strategies were investigated for the input data. The first strategy was to try using only a single time interval, the second was to try simultaneously using several overlapping time intervals and thirdly to try using consecutive equally sized time intervals.

Anticipating that it would be difficult to distinguish between different types of breathing events such as obstructive and central apneas, the network was designed to only distinguish three states: normal breathing, apnea and hypopnea. The three states were represented using three output nodes corresponding to the three states. The output of each node is set to the range 0 to 1. The output state is then taken to be the state corresponding to the output node with output closest to 1. The input data for the network again consisted of the Fourier transform and fractal measure of the data over specified time intervals.

The data from the 9 patients was divided into three groups of three. Each time the experiment was performed two groups were used for training while the third was used for testing. The number of epochs in which a patient is having an event is only a small proportion of the total number of epochs. To make the training data more even the data was sorted so that the amount of non-event training data was comparable to the amount with events. For the single time interval networks the number of nodes in each layer was 10,10,3 while for the multiple time intervals it was 20,20,3.

### 4.3 Results

#### 4.3.1 Single time interval

The experiment was performed using several different sized time intervals. The results are shown in Table 4.1. For each of the three states; normal, hypopnea and apnea, two percentages are given representing how often an occurrence of that state is correctly detected (true positive) and how often other states are correctly scored as being other states (true negative) respectively.

Overall the 15 second interval gives the best performance of the three time intervals. In all cases the true negative accuracy for apneas is very high while the true positive accuracy is poor. This indicates that the networks are reluctant to classify data as an apnea. The true positive accuracy for hypopneas is substantially higher than for apneas and given that the network has a greater tendency to classify data as a hypopnea it is not surprising that the true negative accuracy is less for hypopneas than for apneas.

#### 4.3.2 Overlapping time intervals

The experiment was performed again but this time four overlapping time intervals of lengths 2, 5, 10 and 15 seconds used in producing input data to the network. The results for this network are shown in Table 4.2.

Table 4.1: Network performance for a single time interval.

		normal		hypopnea		apnea	
time	data	positive (%)	negative (%)	positive (%)	negative (%)	positive (%)	negative (%)
15s	train	81.7±6.8	84.3±7.3	80.2±11.4	79.8±6.4	47.8±2.9	95.9±1.6
	test	71.1±17.3	77.5±7.9	70.1±22.9	69.2±5.8	14.8±6.5	93.6±9.7
5s	train	76.8±7.5	86.1±7.1	86.4±6.1	67.1±7.5	18.3±12.5	99.1±0.4
	test	56.1±10.7	54.6±14.3	44.1±14.1	62.6±7.9	5.4±6.1	93.4±8.1
1s	train	83.2±2.4	62.3±9.6	60.1±10.7	79.7±3.4	24.8±5.4	97.6±0.8
	test	62±9.4	43.2±9.5	30.4±2.3	71.5±2.7	18.9±11.6	90.4±8.4

Table 4.2: Network performance for overlapping time intervals.

	normal		hypopnea		apnea	
data	positive (%)	negative (%)	positive (%)	negative (%)	positive (%)	negative (%)
train	82.1±7.0	93.5±1.3	89.2±6.4	85.3±4.8	72.4±4.5	94.4±3.2
test	44.1±11.6	65.6±7.0	39.8±7.6	61.2±2.0	28.6±16.6	79.4±11.0

Compared with the single time interval networks, this network performs better overall on the training data yet overall performs less well on the test data. This may be because the network is being fed with too much redundant information.

#### 4.3.3 Consecutive time intervals

The experiment was performed again this time using four consecutively spaced intervals each of 2 seconds. The results for this network are shown in Table 4.3.

This network is able to fit the training data with very high accuracy but once again on the test data the network performs poorly. Comparing all the different networks reveals that the networks with more input data are more capable of fitting the training data but when it comes to the test data it is actually the single 15 second interval network which performs the best.

Table 4.3: Network performance for consecutive time intervals.

	normal		hypopnea		apnea	
data	positive (%)	negative (%)	positive (%)	negative (%)	positive (%)	negative (%)
train	96.3±3.3	92.4±4.6	92±5.7	94.7±3.4	78.5±7.7	99.1±0.6
test	63.8±5.1	48.9±13.0	29.5±5.9	74.4±3.3	21.9±6.6	85.7±6.5

## 5. CONCLUSIONS

### 5.1 Results

The results of the skin conductivity study indicate that there may be some relation between the skin conductivity and the breathing and heart signals. Despite the larger values of correlation obtained being large enough to suggest some relation between the signals, there is substantial variation between the positions of the correlation peaks from one time to another. This complete lack of consistency suggests that the visual resemblance between the filtered skin conductivity and breathing and heart rate signals is due only the application of a filter centred around the desired frequency rather than due to an influence of breathing or heart beat on skin conductivity. Based on the results of this experiment it does not appear that the heart rate or breathing rate can be obtained from the skin conductivity.

The results of the sleep scoring task show that neural networks are capable of the detection of sleep stages. However the networks poor performance in scoring certain stages, particularly stage 1 and 3 show that more work is required to produce reliable stage scoring. The addition of fractal measure was found to have only a slight improvement on scoring accuracy while equally distributing the training data was found to have a negative effect due to increased presence of the difficult to score stages. There was little difference between the performances of the two chosen networks but

performance was notably better when the network was trained with the active patients indicating that better performance is achieved when the network is trained using the more diverse, irregular training data.

The results of the breathing event detection task show that the EEG does bear some relation to the occurrence of breathing events but that it is difficult to detect breathing events using only the EEG. It was found that hypopneas were consistently easier to detect than apneas but that normal breathing was more often falsely classified as hypopnea than apnea. Of the different time intervals used it was the 15 second time interval that was the most effective even when compared against networks using multiple time intervals.

## 5.2 Future Work

The skin conductivity experiment only used filtering on the skin conductivity. It is possible that more advanced signal processing techniques could produce better results. Another issue is the quality of the signals. The results of this experiment may have turned out more positively if the quality of the signals was superior.

The techniques of Fourier transform and fractal dimension are only two of many possible methods for producing input data for a neural network. One could investigate using windowing to improve the effectiveness of the Fourier transform or one could try using wavelet transforms. There are also many non-linear measures that could be implemented. The choice of neural network node structure was chosen after experimentation. A more effective technique would be to use a genetic algorithm to optimize the node structure<sup>11</sup>. The effects of sampling and the number of quantization bits on the Higuchi method need further investigation.

## REFERENCES

1. Whiteford, L., Fleming, P., Henderson, A. J., "Who should have a sleep study for sleep related breathing disorders?" *Arch Dis Child*, vol. 89, pp. 851-855, 2004.
2. Marcus, C., L., "Sleep-disordered breathing in children," *Current Opinion in Pediatrics*, vol. 12, pp. 208-212, 2000.
3. Huupponen, E., Himanen, S.-L., Hasan, J., Värri, A., "Automatic quantification of light sleep shows differences between apnea patients and healthy subjects", *International Journal of Psychophysiology*, pp. 1-8, 2003.
4. Al-Ani, T., Hamam, Y., Novak, D., Pozzo Mendoza, P., T., Lhotska, L., Lofaso, F., Isabey, D., Fodil, R. 2005, "Noninvasive automatic sleep apnea classification system", *BioMed Sim '05*, Linköping, Sweden.
5. Chervin, R. D., Burns, J. W., Subotic, N. S., Roussi, C., Thelen, B., Ruzicka, D. L., "Correlates of respiratory cycle-related EEG changes in children with sleep-disordered breathing", *Sleep*, vol. 27, pp. 116-121, 2004.
6. Kirby, S., D., Danter, W., George C., F., P., Francovic, T., Ruby, R., R., F., Ferguson, K., A., "Neural network prediction of obstructive sleep apnea from clinical criteria", *Chest*, vol. 116, pp. 409-415, 1999.
7. Family practice notebook.com, [online, accessed 27 Apr. 2005], URL <http://www.fpnotebook.com/LUN54.htm>.
8. Medline Plus, [online, accessed 27 Apr. 2005], URL <http://www.nlm.nih.gov/medlineplus/ency/article/003399.htm>.
9. Accardo, A., Affinito, M., Carrozzi, M., Bouquet, F., "Use of fractal dimension for the analysis of electroencephalographic time series", *Biological Cybernetics*, vol. 77, pp. 339-350, 1997.
10. Riedmiller, M., Braun, H., "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," *Proceedings of the IEEE International Conference on Neural Networks*, San Fransisco, USA, 1993.
11. Koza, J., R., Rice, J., P., "Genetic generation of both the weights and architecture for a neural network", *Proceedings of International Joint Conference on Neural Networks*, Seattle, 1991.