Automatic Detection Of Incidents Of Simultaneous Cardiac And Respiratory Deceleration, And Apnoea In Preterm Infants

Belal SY [1], Emmerson A [2], Beatty PCW [1]
[1] Manchester University, [2] Saint Mary's Hospital, Manchester, UK
Suliman.Belal@manchester.ac.uk

Abstract. The detection of the incidents of simultaneous cardiac and respiratory deceleration, and apnoea in preterm infants is often based on simple threshold techniques, which suffer from poor specificity and are prone to artefacts. Three methods for the automatic detection of such incidents were designed, tested and evaluated from the time series of Heart Rate (HR), Respiratory Rate (RR) and Oxygen Saturation (SpO_2) collected from 54 neonates (approximately 2426 hours of recording). They were the cumulative sum method with shared entropy, the correlation method and the derivative method. The latter method had the highest performance (100% sensitivity, 96.19% specificity and 96.79% accuracy). Though not optimised to work in real time, this method has the potential in forming the basis of a system for detecting incidents of cardiac and respiratory deceleration, and apnoea.

1 Introduction

Apnoea of prematurity (AP) is the cessation or slowing of breathing for a period greater than 15 s associated with falling blood oxygen concentration (desaturation). Apnoeic attacks are common in preterm infants, with the most severe requiring artificial ventilation and drug treatment (1).

The relationship between neonatal apnoea, bradycardia, and desaturation is complex. Levels of Oxygen Saturation (SpO_2) in neonates have been shown to decrease rapidly with apnoea (2), and episodes of bradycardia have been reported to occur in association with apnoea (3). As a result, the persistence of the occurrence of such episodes may compromise oxygenation and tissue perfusion, and medical intervention may be required to prevent severe and permanent damage to the neonate.

Apnoea-like incidents which are associated with HR deceleration and desaturation is a symptom of a wide range of different physiological states. Some of these states are defined as AP, and some are idiopathic. In this research the term apnoea will be used in a general sense.

In the NICU, clinical staff normally detect episodes of apnoea by observing Heart Rate (HR), Respiration Rate (RR) and SpO_2 simultaneously. However, continuous observation is impractical, and attention overload and stress from looking after several patients at once can lead to mistakes by clinical staff (4). Consequently, automatic methods of detecting early apnoea would be advantageous especially if the methods were more reliable and sensitive than the current threshold-based techniques, which have difficulty with coping benign physiological variation (5;6) and measurement limitations.

The aim of the research reported was to develop a method for detecting apnoea using mutual information present in the time series of HR, RR and SpO2 because bradycardia does not occur without respiratory changes, and even the 'non-apnoea' associated bradycardia is always associated with changes in respiratory pattern (7). Such a method would detect apnoea more reliably and takes into account the artefacts in the HR, RR and SpO_2 time series, which might otherwise confused with true apnoeic events.

2 Material and methods

2.1 Subjects and data

HR, RR and SpO_2 from 54 neonates admitted during the period 2000-2004 were retrieved from a database created in the Neonatal Medical Unit, St. Mary's Hospital, Manchester, UK. The data had a sampling rate of 1 Hz and totalled approximately 2426 hours of recording with $\mu = 44h$ and $\sigma = 7h$. The data was anonymised and stored in a way consistent with the UK Data Protection Act and Manchester University's Data Regulations.

2.2 Methods

A combined correlation and derivative detector was implemented to detect the apnoea incidents. This detector was neither tuned nor validated, and will be later described in details. The application of this detector on the data resulted in a set of events, among which might be apnoea events. It is aimed to identify such true events, which can be used in later analysis as training and validation sets. However, because scrutinising and visually inspecting such events is tedious and time consuming, the events were clustered, and the averaged clusters were shown to a clinical expert who identified the clusters presenting true events. As it is not possible to directly cluster the events because they have different durations, several features were extracted from the time series data prior to clustering.

2.2.1. Features extraction

From each event a set of 8 features were extracted; namely: HR-slope, RR-slope, SpO_2-slope, the square of the HR-RR, HR-SpO_2 and RR-SpO_2 correlation coefficients, the sum of the Shannon entropy of HR, RR and SpO_2 during each event, and the duration of the event itself. The slopes of the HR, RR and SpO_2 during the suspected events were calculated using a linear fit model, which was solved using the Weighted Least Absolute Residual (WLAR) method.

2.2.2. Clustering

Because there is no a priori knowledge of the true number of clusters in the features space, the unsupervised k-means clustering algorithm was used to group together suspected apnoea events with similar features. The algorithm was run with different values of *k*, and the Davies–Bouldin (DB) index (8) was calculated for each run. The *k* value which corresponded to the smallest value of the DB index was then chosen, and the corresponding clusters were considered to be the correct solution.

2.2.3. Abrupt changes detection via the cumulative sum (CUSUM) detector

This detector is based on the cumulative sum (CUSUM) approach first introduced in (9). Let $(y_k)_{1\le k\le n}$ be a sequence of independent Gaussian random variables. The aim is to detect the change in the mean value μ of this sequence.

$$\mu_n = \begin{cases} \mu_0 \; if \;\; n < r-1 \\ \mu_1 \; if \;\; n \ge r \end{cases} \tag{1}$$

Where, r is the instant of the change, μ_0 is the mean before the instant of change and μ_1 is the mean after the change. The procedure is then to compute of the cumulative sum for detecting an increase in the mean (10):

$$\begin{aligned} \Lambda_0 &= 0 \\ \Lambda_k &= \sum_{j=1}^{k} (y_j - \mu_0 - \frac{\nu}{2}) \\ \Lambda_{\min} &= \min_{1\le j<k} \Lambda_k \end{aligned} \tag{2}$$

The mean value μ_0 of the analysed series y is calculated over a time window ωl. Λ_k is the likelihood ratio and ν is the reference value, which is the maximum variation that is allowed in the mean value of the signal. An event-detection occurs when the following conditions are satisfied:

$$(\Lambda_k - \Lambda_{\min} \geq \lambda) \cap (r = \arg \min_{1 \leq j < k}(\Lambda_k)) \tag{3}$$

Both ν and λ (the detection threshold) are positive values, which need to be determined.

As the aim of this detector is to classify each value of the sequence y to 'event' or 'no event' , the area under the receiver operator characteristics curve (ROC) curve (A_{ROC}) can be calculated given that an expert-validated data set has already been constructed. Hence, the free variables can be adjusted as to maximise A_{ROC} .

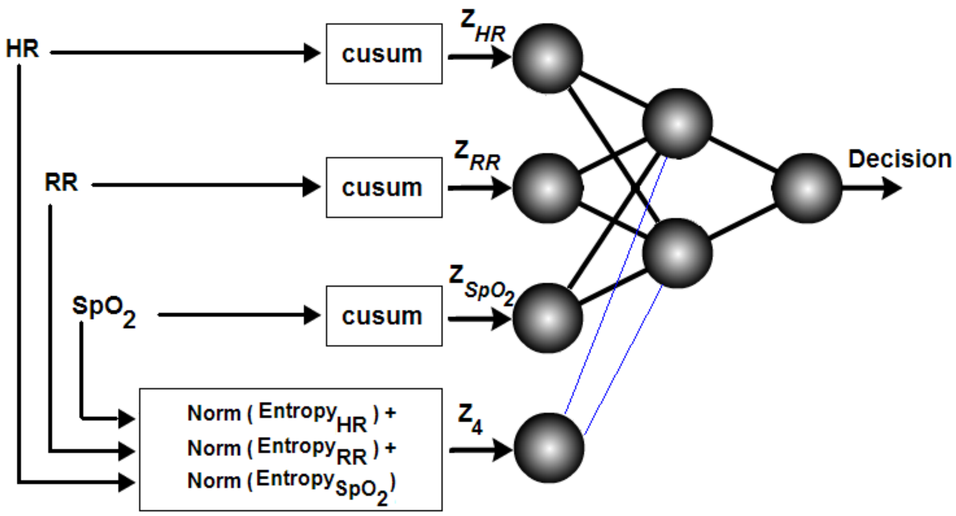

Fig 1.A feed-forward neural network with four inputs: the three outputs of the CUSUM detectors and the sum of normalised entropies of the HR, RR and SpO_2 time series.

The best performance can be achieved by finding the value of the free parameters through the following optimisation:

$$(\hat{\omega} l_{HR}, \hat{\nu}_{HR}, \hat{\lambda}_{HR}) = \arg \min_{a_1 \leq \omega l \leq b_1,\, a_2 \leq \nu \leq b_2,\, a_3 \leq \lambda \leq b_3} (-A_{ROC}(z)) \tag{4}$$

The above multi-dimensional minimisation was solved using the downhill Simplex method (11). Because the Simplex algorithm is sensitive to the initial conditions, an extensive set of test runs showed that the best results were obtained when the algorithm is restarted several times until this does not result in a further improvement. This method was also applied on the RR and SpO_2 time series, and their corresponding free parameters were calculated. Next, the sum of the normalised Shannon entropy of the three time series y_j of HR, RR and SpO_2 was calculated in a moving window of length n = 60 seconds according to the following equation:

$$z_4 = \sum_{j=1}^{j=3} norm(\sum_{i=1}^{i=n} {y_j}^2(i) \log({y_j}^2(i))) \tag{5}$$

Where, $Norm$ is a normalisation function defined as:

$$Norm(y) = \frac{y}{y_{\max} - y_{\min}} \tag{6}$$

Because the temporal relationship between apnoea, desaturation and bradycardia is not clear since various hypotheses indicated different orders of sequence of events. Thus, it is possible to treat the mathematical model governing this temporal relationship as a black-box. Artificial Neural Networks (ANN) are, hence, suited to represent this model. Because the true class (ground-truth) for labelling the data is known via the clinical expert input, supervised neural networks are the choice for this application. Therefore, a feed-forward multilayer-perceptron (MLP) neural network with four inputs (z_{HR}, z_{RR}, z_{SpO_2}, and z_4) and one output was constructed, trained and its performance was calculated, see Fig 1. In order to decide upon the number of layers needed, types of transfer functions, number of neurons in the hidden layer(s), the learning rate and the initial weights effect on the network performance, various networks were systematically created and trained and their performance was calculated.

2.2.4. The correlation-based detector

The aim of this detector is to quantify each pair of correlations between HR, RR and SpO_2 at every instant of time (every one second in this data set), then use this quantification to discriminate between the instances that constitute 'event' or 'no event'.

The correlation coefficients of each pair of parameters were calculated in a running window of 60-seconds, but this window could have been chosen systematically via optimisation.

A neural network with three inputs: the three correlation coefficients of HR, RR and SpO_2 was designed. The performance of this neural network was calculated with and without Principal Components Analysis (PCA) as a pre-processing step.

2.2.5. The derivative-based detector

The instantaneous derivative $f(t)$ of each data sample of the three parameters, HR, RR and SpO_2 was calculated. A similar approach was reported in (4). For a given time series x of length n, the derivative f is given by:

$$f(i) = \frac{1}{2dt}(x_{i+1} - x_{i-1});\ where\ i = 0,1,2,\ldots n-1 \qquad (7)$$

The instantaneous values of each resulting function were compared with a threshold δ_{grad} chosen to maximise A_{ROC} after having been joined with the AND operator and been median-filtered (equation (9)).

$$\hat{\delta} = \arg \max_{a \le \delta \le b, c \le w \le d} A_{ROC} \left\{ median_w \left(\prod_{j=1}^{j=3} \delta \frac{d}{dt} x_j \right) \right\} \qquad (8)$$

A median filter of length w was used to smooth the output. Consequently, this would increase the detector specificity and make it less sensitive to artefacts.

3 Results

The initial combined derivative and correlation detector was found to contain 104 episodes of suspected apnoea events. The events durations had a mean of $\mu = 88.33\,s$ and a standard deviation $\sigma = 26.91\,s$ and a range of 134 s. Clustering revealed 7 clusters, and the HR, RR and SpO_2 time series corresponding to each event in each cluster were averaged to form a cluster prototype (Fig 2.).

Fig 2. was shown to a clinical expert, who identified clusters 1-6 as true events. Hence, the data from these clusters were considered the main data set which will be used for the remaining of this research. The identified events (98 events: 54880 s in which 8397 s identified as apnoea) were divided into two equal parts: the detector training data set and the detector evaluation data set. The detector training set is further divided into neural networks (training, validation, testing) data sets with ratios of 50%, 25%, 25%, respectively.

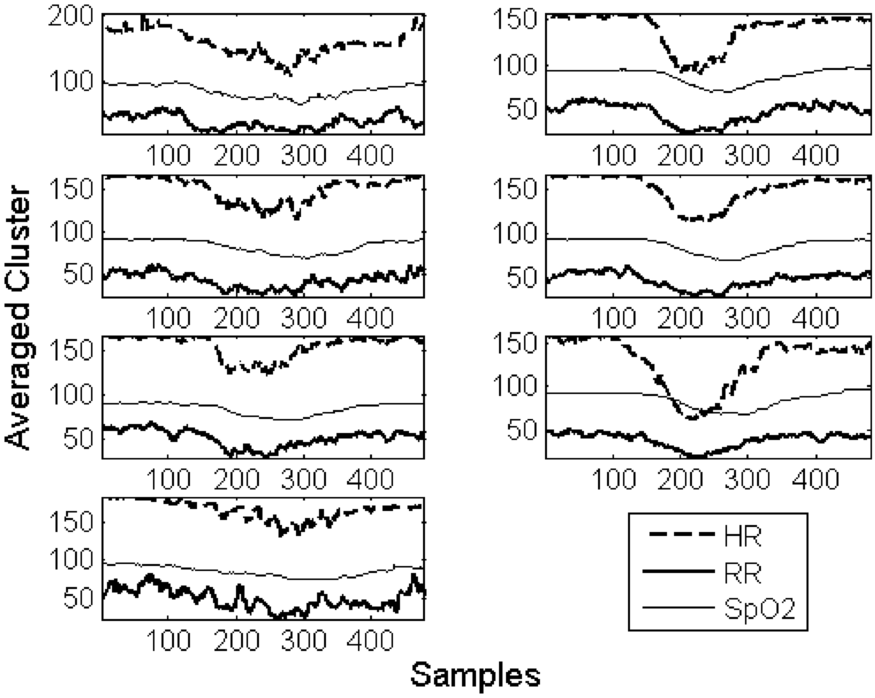

Fig 2.The averages of the events in each cluster.

The values of the optimised free parameters of the CUSUM-based detectors and the performance obtained as a result of applying the individual detectors on the detector evaluation data set are presented in Tab 1.

Detector	A_{ROC} %	**Sensitivity** %	**Specificity** %	**Accuracy** %	ωl	λ	ν
HR	75.11	81.20	69.03	70.94	36	27.6370	2.0845
RR	65.27	69.27	61.28	62.54	45	10.1860	0.0875
SpO_2	65.96	86.90	45.01	51.61	56	32.0108	1.2377

Tab 1. The performance of the individual CUSUM-based detectors using the detector evaluation data set. Note that the free parameters were calculated via optimising the detectors using the detector training data sets.

The ROC curves shown in Fig 3. were obtained when applying the individual detectors on the detector evaluation data set.

A multi-layer feed forward neural network with four inputs (z_{HR}, z_{RR}, z_{SpO_2} and z_4) and one output was constructed. The Back-propagation algorithm: Gradient descent with momentum was chosen for training. In order to improve generalisation, early stopping during learning is used.

Next, in order to decide the network which gives the most robust performance, the number of layers, neurons in the hidden layer(s) and the types of transfer functions were varied and studied systematically. A neural network with one hidden layer of two neurons was chosen. The transfer functions for the hidden layer were decided to be log-sigmoid and the transfer function for the output layer was decided to be a tan-sigmoid. The resulting ANN-CUSUM based detector was evaluated and the performance shown in Tab 2. was obtained.

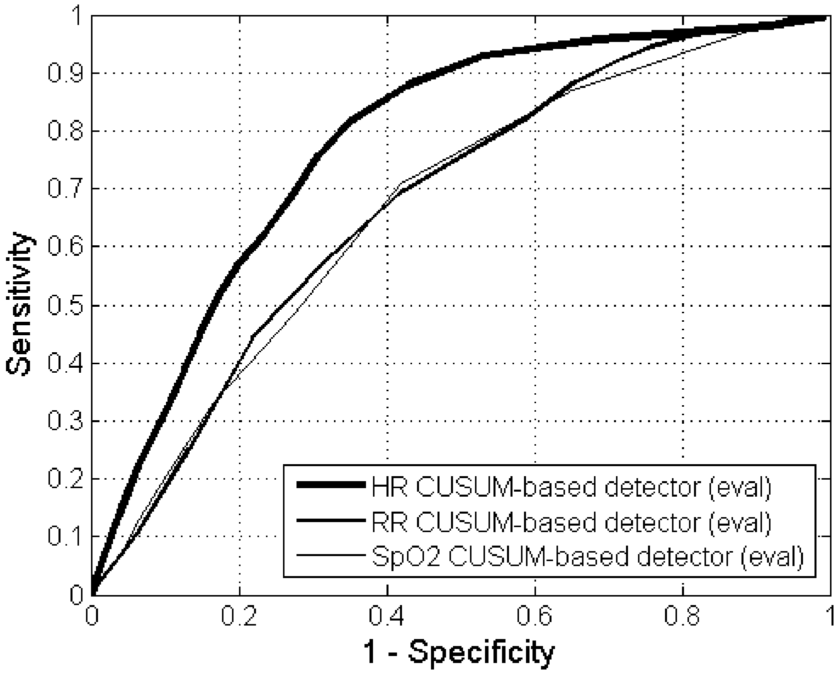

Fig 3.The ROC curves for the individual CUSUM-based detectors for HR, RR and SpO_2 obtained from the detector evaluation data set.

Data Set	A_{ROC} %	Sensitivity %	Specificity %	Accuracy %	Threshold
Training	85.06	91.67	78.45	85.06	0.4132
Validation	84.99	91.59	78.38	84.99	
Testing	84.76	91.44	78.08	84.76	
Evaluation	84.63	94.53	74.72	77.84	

Tab 2. The performance of the chosen ANN-CUSUM and entropy based detector using the ANN training, validation and testing data along with the detector evaluation data set.

The neural networks programmes were written in Matlab 6.5 (The MathWorks R13) running on a Windows XP Operating System.

For the correlation-based detector, the neural network had three inputs, one hidden layer with 2 neurons and one output neuron. The performance of this detector was found to be 88.79% sensitivity, 72.03% specificity and 74.68% accuracy.

The neural network training data was normalised to have zero mean and unit variance. PCA was then applied on the input space retaining only those components that contribute

[55% to 100%] to the variance in the data set with a step of 5%. By doing this, the dimensions of the input space are reduced and variance is maximised. Each time after the application of PCA, a feed forward MLP network was designed with a number of inputs equivalent to the number of PCA dimensions, one hidden layer (with 2 neurons) and one output layer. The maximum number of allowed epochs was chosen to be 7000 epochs, the MSE goal was 0.001, and the momentum constant was 0.2. The learning rate was 0.3.

The data, as a result, was projected on the set of principal components accounting for 70% of the variance, which gave the highest performance in terms of specificity. The performance of the chosen ANN was 81.85% sensitivity, 75.83% specificity and 76.78% accuracy.

For the derivative-based detector, the threshold which was obtained by maximising A_{ROC} is presented in Tab 3. along with the detector's performance.

Detector	A_{ROC} %	**Sensitivity %**	**Specificity %**	**Accuracy %**	w	δ_{grad}
With median	98.09	100.00	96.19	96.79	56	-0.0292
Without median	95.11	95.53	94.68	94.82	n/a	

Tab 3. The performance of the derivative-based detectors using the detector evaluation data.

4 Conclusions

Using clustering served the purpose of providing basis for the clinical expert to identify true physiological events, and it fulfilled a task which otherwise is time consuming.

Three types of detectors were investigated: cumulative sum-based, correlation-based and derivative-based detectors.

The HR CUSUM-based detector showed reasonable performance in terms of sensitivity (81.20%), but the specificity was low (69.03%). It can be induced that changes in the behaviour or the characteristics of the HR time series during the events of interest is more persistent/cohesive than the changes in the characteristics of the RR and SpO_2 time series, if studied similarly on their own.

The optimised HR CUSUM-based detector showed that changes of the mean heart rate of $\nu \geq 2$ beats per $wl = 36$ seconds are indicative of events of cardiac deceleration which are likely to be associated with simultaneous respiratory deceleration and apnoea.

The events detection can be improved if the system utilised the cumulative sum of HR, RR and SpO_2 in addition to the shared entropy of the three time series. If those were used in an artificial neural network, both sensitivity and specificity can be improved.

Using a combined PCA and neural network approach on the time series correlation coefficients, a slight an improved specificity was observed (75.83%), but on the account of sensitivity (81.85%). On the other hand, the derivative-based with median detector gave the best detection performance among all the proposed classifiers.

The processes developed in this project can be used successfully to extract a set of incidents of simultaneous cardiac and respiratory deceleration, and apnoea in preterm infants. This set can be further analysed to study the nature of such events and perhaps their temporal relationships, which will enhance our understanding to their nature.

References

[1] Martin RJ, bu-Shaweesh JM, Baird TM. Apnoea of prematurity. Paediatr Respir Rev 2004; 5 Suppl A:S377-S382.
[2] Henderson-Smart DJ. Vulnerability to hypoxemia in the newborn. Sleep 1980; 3:331-42.
[3] Girling DJ. Changes in heart rate, blood pressure, and pulse pressure during apnoeic attacks in newborn babies. Arch Dis Child 1972; 47:405-10.
[4] Portet F, Feng G, Hunter J, Sripada S. Evaluation of On-Line Bradycardia Boundary Detectors from Neonatal Clinical Data. 2007 p. 3288-91.
[5] Southall DP, Johnston F, Shinebourne EA, Johnston PG. 24-hour electrocardiographic study of heart rate and rhythm patterns in population of healthy children. Br Heart J 1981; 45:281-91.
[6] Davis N, Sweeney LB. Infantile apnea monitoring and SIDS. J Pediatr Health Care 1989; 3:76-5.
[7] Smith ML, Milner AD. Bradycardia and associated respiratory changes in neonates. Arch Dis Child 1981; 56:645-8.
[8] Davies DL, Bouldin DW. A Cluster Separation Measure. Pattern Analysis and Machine Intelligence, IEEE Transactions on 1979; PAMI-1:224-7.
[9] Hinkley DV. Inference about the change-point in a sequence of random variables. Biometrika 1970; 57:1-17.
[10] Basseville M, Nikiforov I. Detection of abrupts changes: Theory and applications. Prentice Hall; 1993.
[11] Nelder JA, Mead R. A simplex method for function minimization. Computer Journal 1965; 7:308-13.