ARTIFACT DETECTION IN THE PO_2 AND PCO_2 TIME SERIES MONITORING DATA FROM PRETERM INFANTS

Cungen Cao,[1] PhD, Neil McIntosh,[2] DSc(Med), Isaac S. Kohane, MD, PhD, and Kang Wang, PhD

Cao C, McIntosh N, Kohane IS, Wang K. Artifact detection in the PO_2 and PCO_2 time series monitoring data from preterm infants.
J Clin Monit 1999; 15: 369–378

ABSTRACT. Background. Artifacts in clinical intensive care monitoring lead to false alarms and complicate later data analysis. Artifacts must be identified and processed to obtain clear information. In this paper, we present a method for detecting artifacts in PCO_2 and PO_2 physiological monitoring data from preterm infants. **Patients and data.** Monitored PO_2 and PCO_2 data (1 value per minute) from 10 preterm infants requiring intensive care were used for these experiments. A domain expert was used to review and confirm the detected artifact. **Methods.** Three different classes of artifact detectors (i.e., limit-based detectors, deviation-based detectors, and correlation-based detectors) were designed and used. Each identified artifacts from a different perspective. Integrating the individual detectors, we developed a parametric artifact detector, called ArtiDetect. By an exhaustive search in the space of ArtiDetect instances, we successfully discovered an optimal instance, denoted as ArtiDetector. **Results.** The sensitivity and specificity of ArtiDetector for PO_2 artifacts is 95.0% (SD = 4.5%) and 94.2% (SD = 4.5%), respectively. The sensitivity and specificity of ArtiDetector for PCO_2 artifacts is 97.2% (SD = 3.6%) and 94.1% (SD = 4.2%), respectively. Moreover, 97.0% and 98.0% of the artifactual episodes in the PO_2 and PCO_2 channels respectively are confirmed by ArtiDetector. **Conclusions.** Based on the judgement of the expert, our detection method detects most PO_2 and PCO_2 artifacts and artifactual episodes in the 10 randomly selected preterm infants. The method makes little use of domain knowledge, and can be easily extended to detect artifacts in other monitoring channels.

KEY WORDS. Monitoring, preterm infant, physiological time series data, artifact detection, artifactual correlation, artifactual episode.

INTRODUCTION

The data generated by ICU monitors are potentially valuable in detecting physiological trends and pathological diagnoses in critically ill patients. It is probable that early warning of developing problems will lead to more timely intervention with reduction in mortality and morbidity. However, artifacts are common in the data. These often prevent the identification of important events, and reduce the trust placed by staff on the machines [1, 2]. Frequent false alarms distract clinical staff and may frighten patients and relatives. Actions resulting from artifactual data may be unnecessary or inappropriate, and appropriate action may not be initiated if an important event is missed. Artifact identification is important both for visual interpretation by an observer and as a basis for more automatic decision support [3–9].

From the Informatics Program, Children's Hospital, Harvard Medical School, 300 Longwood Avenue, Boston, MA 02115, U.S.A.

Received Mar 30, 1999, and in revised form Aug 11, 1999. Accepted for publication Aug 19, 1999.

Address correspondence to Dr Cungen Cao, Informatics Program, Children's Hospital, 300 Longwood Avenue, Boston, MA 02115, U.S.A.

[1] A portion of the work was done when this author visited the Clinical Decision Making Group, Laboratory for Computer Science, MIT.
[2] On sabbatical from the University of Edinburgh.

Journal of Clinical Monitoring and Computing **15**: 369–378, 1999.

Artifact detection based on domain knowledge is powerful [6]. However, "knowledge" may not be available in some domains. For example, in monitoring preterm infants, we do not know all the event patterns that cause artifacts. When monitoring physiological parameters (e.g., the PCO_2 and PO_2), clinical examination and investigation can often be seen to disturb the infant. This in turn disturbs to a variable extent the monitor readings. PO_2/PCO_2 probe repositions usually cause consistent changes: The PO_2 rises to about 20 kPa, and the PCO_2 drops close to zero as the probe is removed from the infant – these are the levels of oxygen and carbon dioxide in the air. Other events, such as changing diapers and the infant crying or spontaneously moving, may also change the PO_2 and PCO_2 readings, though in a less consistent way.

Using offline experiments and manual data analysis, Cunningham et al. concluded that artifact identification largely depends on the investigator's personal understanding of the data [4]. When clinicians agree on what artifacts are, their detection rates are consistent from investigator to investigator. Offline or retrospective artifact detection is also made difficult by the poor documentation that usually accompanies monitoring. When a clinician examines an infant, artifacts may be generated in a number of monitoring channels (e.g., the heart rate and blood pressure). It is unusual for the exact time of the examination to be noted in conjunction with the monitoring data, particularly if the examination is made in an emergency situation. Events may be noted in the monitoring system at some later time, after the completion of a procedure, but often they will not be noted at all. Events such as an infant's spontaneous movements may be unnoticed or ignored.

In this paper, we are interested in the development of detection methods without considering data annotations. Sittig and Factor used a Kalman filtering technique to automatically identify artifacts in cardiovascular monitoring. Although the technique performed well, it was often difficult to set the model parameters and variable covariances in a systematic manner [5]. In 1998, we reported a simple method for detecting artifacts in a single data stream that required little domain knowledge [3]. The method derived two new data streams from a single original stream by comparing two successive values in the original stream. Based on the two derived streams, linear regression lines and nonlinear regression curves predicted further values in the original data stream. If an observed value significantly deviated from its predicted value, the observed value was likely to be an artifact. Experiments showed that the method detected most artifacts (99.0%) in a single data stream from a well infant, but could miss a substantial proportion of the artifacts from an ill infant (up to 25%). In considering only a single data stream, this method could not utilise correlated artifacts between multiple channels.

This paper develops a new method of detecting artifacts in combined PO_2 and PCO_2 physiological time series data from preterm infants.

METHODS

Clinical database, training dataset and gold standard

Clinically, each infant is monitored by several devices, giving multiple channels of physiological information. In this experimental work, we considered two important channels, the PO_2 and PCO_2, which mainly monitor the respiratory system.

To train and discover optimal detectors, we randomly chose 10 preterm infants from a database of 153 high-risk infants receiving intensive care. The monitors used in our neonatal unit are Hewlett Packard 78344 multichannel neonatal monitors the data from which are sampled to computer 1/second, and one-minute averages of this data are saved to the database. In this study, we randomly chose a 10-hour data segment from the original data streams of each infant. This gave 600 records or pairs of PO_2 and PCO_2 values for each selected infant, and 6000 records of PO_2 and PCO_2 values in total.

The gold standard against which the artifact detection method was compared was a domain expert (clinical neonatologist – NM). The expert examined the original trend graph data together with its incomplete annotations, independently identifying and noting each one-minute data value which was more or less than was appropriate for the infant.

Developing an artifact detection method

From an understanding of artifacts in physiological time series data, we have designed three types of artifact detectors. These detectors are used in combination in an attempt to identify artifacts in each of the PO_2 and PCO_2 channels from three different perspectives. Based on these three artifact detectors, we have developed a parametric artifact detector, called ArtiDetect. When specific values are assigned for the parameters in ArtiDetect, a specific ArtiDetect instance is determined. A search is then made in the space of ArtiDetect instances for optimal instances.

PO2 stream	... 8.5 8.7 8.7 8.7 8.6 8.7 **14.1** **15.2** **8.9** 8.8 8.7 8.5 ...
PCO2 stream	... 4.2 4.3 4.3 4.3 4.3 4.2 **3.2** **3.8** 4.3 4.2 4.3 4.3 ...
	→ Time (min)

Fig. 1. Illustration of multi-channel data streams and artifacts (in bold face).

Limit-based detectors

In clinical medicine, the limits of various physiological parameters play an important role in determining whether patients are normal or abnormal. In Figure 1, if we assume that the upper limit of PO_2 is 14.8 kPa,* the PO_2 value of 15.2 kPa is immediately identified as a PO_2 artifact.

We denote the lower and upper limits of the limit-based detector for PO_2 artifacts as *lpo2* and *upo2*, and the lower and upper limit of the limit-based detector for PCO_2 artifacts as *lpco2* and *upco2*, respectively.

Deviation-based detectors

When wide limits are adopted for the limit-based detectors, the detectors may have a high true positive rate (sensitivity) but an unacceptable false positive rate (lack of specificity). If the limits are reduced, the sensitivity may decrease though with better specificity. To improve the accuracy of the limit-based detectors, we added in deviation-based detectors. The deviation-based detectors monitor rapid changes (deviations) in physiological parameters that are more rapid than would be possible in the infant.

The deviation-based detector for PO_2 works as follows. First, it has a moving time window of length t. At this point we do not know what length is appropriate, and therefore we leave it as a parameter to be determined by experiments in the next section. Second, if the standard deviation of the PO_2 values in a moving time window is beyond a threshold *dpo2*, some value within the data window is considered to be an artifact. Moreover, if we know that the first $t-1$ values in the window are not artifacts, the last value must be the artifact. That is, it is the last value that causes the standard deviation of all the values in the moving window to be higher than *dpo2* that is labelled as an artifact.

The deviation-based detector for PCO_2 functions similarly. First, it has a moving time window of length t'. In this paper, we assume that t' is equal to t. This assumption greatly reduces the search complexity. Further, the assumption is reasonable, because the PCO_2 and PO_2 probe in our unit is a combined device. Second, if the standard deviation of the PCO_2 values within a moving time window is beyond a threshold *dpco2*, some value within the data window is likely to be an artifact. If we know that the first $t-1$ values in the moving window are not artifacts, we claim that the last value is an artifact.

This is illustrated in Figure 1. If we assume that the upper limit of PO_2 is 14.8, then 14.1 can not be identified as an artifact by the limit-based detector associated with the PO_2 channel. However, if we assume that the length of a moving data window to be 6, and set the window to contain the 6 PO_2 values 8.7, 8.7, 8.7, 8.6, 8.7 and 14.1, then the standard deviation of those six values is 2.21. If we assume *dpo2* to be 0.8 in the deviation-based detector for PO_2, 14.1 is immediately identified as an artifact, since the other values in the moving window, i.e., 8.7, 8.7, 8.7, 8.6, and 8.7, are not identified either by the limit-based or deviation-based detector as artifacts.

Correlation-based detectors

When monitoring multiple data channels, artifacts in one channel can imply artifacts in another. Such artifactual correlation may help identify artifacts missed by limit-based or deviation-based detectors.[†] With our combined PO_2/PCO_2 probe PO_2 artifacts are usually mirrored by PCO_2 artifacts and vice versa.

We designed a correlation-based detector for each of the PO_2 and PCO_2 channels. The correlation-based detector for PO_2 artifacts works as follows. If an artifact is detected in the PO_2 channel (either by its limit-based detector or deviation-based detector), the correlation-based detector for PO_2 artifacts is invoked to check if the corresponding value in the PCO_2 channel has a standard deviation greater than a threshold called *cpco2*. If so, the corresponding value in PCO_2 is also considered

* In this paper, the units for PO_2 and PCO_2 are kPa. We will omit the units in the remainder of the paper.

[†] Artifactual correlation between two channels does not necessarily imply that the two channels correlate with each other in normal situations, two channels may correlate with each other only when artifacts occur.

artifactual.* The correlation-based detector for PCO_2 artifacts works similarly to give a standard deviation greater than *cpo2*. Note that *dpo2* is generally greater than *cpo2*, and *dpco2* is greater than *cpco2*; otherwise, correlation-based detectors would be unnecessary, because deviation-based and limit-based detectors would suffice.

In Figure 1, if we assume the length of the moving data window to be 6, the window to contain 8.7, 8.7, 8.7, 8.6, 8.7 and 14.1, and the deviation threshold in the deviation-based detector for PO_2 to be 1.01, 14.1 is identified as an artifact because it is 14.1 that causes the standard deviation to exceed the threshold. The question is now whether the corresponding value of PCO_2 (i.e., 3.2) is also an artifact. In the PCO_2 channel, the standard deviation within the same moving data window is 0.44. If we assume the threshold of the correlation-based detector of PO_2, i.e., *cpco2*, to be 0.33, then 3.2 is also artifactual.

Correcting for artifacts

When artifacts are identified, the artifactual values must be adjusted in order to continue the artifact detection process. This could be performed by a sophisticated technique, e.g. Kalman filtering. However, we have elected to use a simple heuristic rule as follows. Suppose that the length of the moving data window is t, and the window has the t PO_2 (or PCO_2) values V1, V2, ..., Vt. Assume that the first $t-1$ values are not artifacts or have been processed if some of them are artifacts. Our rule says that if Vt is detected by the limited-based detector of PO_2 (or PCO_2), we use the mean of the first $t-1$ PO_2 (or PCO_2) values as an approximation for the artifactual value Vt; otherwise, if Vt is detected by the deviation-based detector or correlation-based detector for PO_2, we use $[3\times(\mathrm{V}1+\mathrm{V}2+\ldots+\mathrm{V}t-1)+\mathrm{V}t]/4$ as an approximation for it.

The rationale for this rule is that a PO_2 (or PCO_2) value tells little if it is beyond the low and upper limits of the limit-based detector for PO_2 (or PCO_2), and therefore we should not use it in calculating an approximation for it. Instead, we use the mean of the previous $t-1$ values within a moving window as a substitute. In the second part of the rule, we use the artifactual value in calculating its substitute, because it may contain some useful information.† Nevertheless, it is accepted that the $t-1$ values prior to Vt should be more reliable than Vt. That is why V1, V2, ..., V$t-1$ have a higher weight (i.e., 3) in calculating the approximation for Vt.

* As the PCO_2/PO_2 probe used is a combined device, we assume that there is no time lag in the correlated artifacts in the PO_2 and PCO_2 channels: once an artifact is detected in PO_2 (or PCO_2) at time t, we immediately check whether there is an artifact in PCO_2 (or PO_2) at the same time t.

† This decision prevents the detectors from locking up in a mode where they regard the true values as artifacts after a long series of artifacts emerges.

ArtiDetect: A parametric artifact detector for PO_2 and PCO_2 artifacts

Based on the individual artifact detectors for channels PO_2 and PCO_2, we designed a parametric artifact detector, called ArtiDetect, for detecting artifacts in PO_2 and PCO_2 monitoring data streams. ArtiDetect is parametric because its component artifact detectors involve several parameters and its overall performance depends on the specific values chosen for these parameters.

ArtiDetect consists of one limit-based, deviation-based, and correlation-based detector for each of the PO_2 and PCO_2 channels. The logic of ArtiDetect is as follows.

Any PO_2 value which is indicated by the limit-based detector or deviation-based detector of PO_2 as an artifact is considered to be a PO_2 artifact by ArtiDetect.

If a PO_2 value is an artifact, and the corresponding PCO_2 value causes a standard deviation greater than *cpco2*, then the corresponding PCO_2 value is also an artifact (in the PCO_2 channel) according to the correlation-based detector associated with PO_2. ArtiDetect then also considers that PCO_2 value to be an artifact.

Any PCO_2 value which is indicated by the limit-based detector or deviation-based detector of PCO_2 is an artifact in the PCO_2 channel is also considered to be an artifact by ArtiDetect.

If a PCO_2 value is an artifact, and the corresponding PO_2 value causes a standard deviation greater than *cpo2*, then the corresponding PO_2 value is also an artifact (in the PO_2 channel) according to the correlation-based detector of PCO_2. ArtiDetect then also considers that PO_2 value to be an artifact.

ArtiDetect instances and their performance

Given specific values for *lpo2, upo2, lpco2, upco2, dpo2, dpco2, cpco2, cpo2*, and *t*, we obtain an ArtiDetect instance, denoted as ArtiDetect(*lpo2, upo2, lpco2, upco2, dpo2, dpco2, cpco2, cpo2, t*). Automatically, all ArtiDetect instances inherit the logic from the parametric Arti-

Detect. Therefore, all ArtiDetect instances are actually artifact detectors for both the PO_2 and PCO_2 channels, though only some of them will have an acceptable performance.

When an ArtiDetect instance is run on a training dataset of N infants, we obtain the sensitivity and specificity of the ArtiDetect instance for each channel in each training infant. Let sens(PO_2) represent the mean of the N individual sensitivity values for the N infants, with a standard deviation derived accordingly. Similarly, let spec(PO_2) stand for the mean of the N individual specificity values for the N infants, also with a standard deviation derived accordingly.

Different ArtiDetect instances have different sensitivity and specificity for each infant and the PO_2 and PCO_2 channels. An optimality criterion for determining which ArtiDetect instances are optimal must be defined. In this paper, we define a criterion based on error minimisation: we look for those ArtiDetect instances where the *error* = [1−sens(PO_2)] + [1−spec(PO_2)] + [1−sens(PCO_2)] + [1−spec(PCO_2)] is minimal, i.e., those instances where the sum of their false positive rates and false negative rates in the PO_2 and PCO_2 channels is minimal.

This criterion is fair in the sense that the performance of artifact detection for a particular channel is not emphasised. However, for different applications different criteria might apply. For example, if false negatives in oxygenation were to be avoided, one might search for an ArtiDetect instance with a maximal sensitivity in the PO_2 channel. Such a criterion could be defined by a joint condition: the sensitivity in the PO_2 channel is maximal and [1−spec(PO_2)] + [1−sens(PCO_2)] + [1−spec(PCO_2)] is minimal.

EXPERIMENTS AND RESULTS

Important centiles of PO_2 and PCO_2

The centiles of PO_2 and PCO_2 were computed from our clinical database of 153 preterm infants. Some of the important centiles are given in Table 1.

Table 1. Important centiles

Centile	PO_2	PCO_2
2nd	0.6	0.1
2.5th	2.5	0.15
3rd	3.5	0.2
5th	4.8	1.0
6th	5.1	1.3
7th	5.3	1.7
10th	5.9	2.3
20th	6.8	3.4
50th	8.3	4.5
90th	11.7	6.8
95th	13.6	7.8
96th	14.3	8.1
96.5th	14.8	8.3
97th	16.3	8.6
98th	19.3	9.3
99th	20.6	10.3
100th	34.0	17.0

Some statistics of the training dataset

The expert neonatologist examined the PO_2 and PCO_2 data stream values in conjunction with any associated annotations. Table 2 shows the numbers of artifactual values in the PO_2 channel (#PO_2) and PCO_2 channel (#PCO_2) for each infant. The table also includes the number of artifactual episodes, defined as a continuous segment of artifacts in a channel. The #Epo2 and #Epco2 fields denote the numbers of the artifactual episodes in PO_2 and PCO_2 channels, respectively.

Table 2. Artifacts identified by expert

	#PO_2	#PCO_2	#Epo2	#Epco2
Infant 1	29	24	5	3
Infant 2	54	41	9	5
Infant 3	43	40	7	7
Infant 4	75	104	3	3
Infant 5	26	22	3	3
Infant 6	120	100	18	12
Infant 7	58	55	4	5
Infant 8	164	124	3	6
Infant 9	24	21	4	2
Infant 10	19	36	4	13
Total	612	567	50	59

From the artifacts identified by the expert in the training dataset, it was found that the probability of a PO_2 value being an artifact, given that the corresponding PCO_2 value is an artifact, is 77.27%. The probability of a PCO_2 value being an artifact, given that the corresponding PO_2 value is an artifact, is 83.95%.

Defining the search space of ArtiDetect instances

For each value assignment for the parameters in ArtiDetect, we obtain an ArtiDetect instance, which has a sensitivity and specificity when run on a training infant.

However, it is generally hard to find out the subset of ArtiDetect instances that have both high sensitivity and specificity. The reason is two-fold. First, there are many parameters to and consequently a very large instance space to explore. Second, all the parameters may interact with each other in a complex manner. For example, the length of moving time window (i.e., t) will affect *dpo2, dpco2, cpco2*, and *cpo2*. Therefore, we relied on an exhaustive search for optimal ArtiDetect instances in the space of ArtiDetect instances.

Potentially, the search space of ArtiDetect instances is infinite., We employed several heuristics to reduce the space while maximising the possibility that all optimal ArtiDetect instances would be included in the space. We defined the search space first by determining a fine-grained set of possible values for each parameter. In the process of determination, we made use of the centiles of PO_2 and PCO_2.

For t:

- We set the possible value set of t to be greater than 1 and less than 11, i.e., artifacts in both PO_2 and PCO_2 channels should be detected with a moving time window of a length greater than 1 but less than 11. This heuristic was supported by our preliminary experiments [e.g., 3].

For the PO_2 channel:

- We chose 1.9 and 14.8 as the lower and upper limits for the component limit-based detector of ArtiDetect, respectively. (1.9 is less the 2.5th centile of PO_2 and 14.8 is the 96.5th). A PO_2 value out of [1.9, 14.8] should have a very high probability of being an artifact. We tried other PO_2 centiles greater than the 96.5th centile (i.e., 14.8) for the upper limit of PO_2. In the experiments, all these different centiles produce the same optimal ArtiDetect instances as the 96.5th centile does.
- For the deviation-based detector of PO_2 in ArtiDetect, we chose the possible value set of *dpo2* to be the set $DPO2 = \{d|d = I \times 0.01$, where I is any integer in $[0, 3000]\}$. Note that the least and largest values in that set is 0.01 and 30.0, respectively. A simple calculation with our clinical database of 153 infants showed that, given any $t \in [2, 10]$, the probability that the standard deviation of the t PO_2 values in a moving time window was less than 30.0 is 99.99%.

For the PCO_2 channel:

- We chose 1.7 and 16.9 as the lower and upper limits for the component limit-based detector of ArtiDetect, respectively. (1.7 is the 7th centile of PCO_2, and 17.0 is the largest value of PCO_2 in our clinical database of the 153 infants that are actually artifacts). We tried some other centiles smaller than the 7th centile for the lower limit of PCO_2, but found that they produced the same optimal ArtiDetect instances as the 7th PCO_2 centile (i.e., 1.7).
- For the deviation-based detector of PCO_2 in ArtiDetect, we set the possible value set of *dpco2* to be the set $DPCO2 = \{d|d = 0.01 \times I$, where I is any integer in $[0, 1700]\}$. Note that the least and largest values in the set were 0.01 and 17.0, respectively. Therefore, *DPCO2* was sufficiently large. In fact, a simple calculation with our clinical database of 153 infants showed that, given any $t \in [2, 10]$, the standard deviation of the t values of PCO_2 in any moving time window was less than 17.0.

Finally, for the correlation-based detectors of PO_2 and PCO_2 in ArtiDetect:

- Note that *dpco2* was generally greater than *cpco2*; otherwise, correlation-based detectors in ArtiDetect would be unnecessary, because deviation-based and limit-based detectors would suffice. For the correlation-based detector of PO_2 in ArtiDetect, we chose the possible value set of *cpco2* to be the set $CPCO2 = \{c|c = dpco2 \times P$, where $dpco2 \in DPCO2$ and P is any one of $0.1, 0.2, 0.3, 0.4, \ldots, 1.0\}$. Note that *CPCO2* was 10 times bigger than *DPCO2*. One could make *CPCO2* finer by letting P take more fractions, such as 0.15 and 0.25. But this would have made *CPCO2* too large as would be the resulting search overhead. On the other hand, the experimental results showed that *CPCO2* was fine enough to find the optimal ArtiDetect instances.
- Following the same argument as the above, we chose the possible value set of *cpo2* to be the set $CPO2 = \{c|c = dpo2 \times P$, where $dpo2 \in DPO2$ and P is any one of $0.1, 0.2, 0.3, 0.4, \ldots, 1.0\}$.

Since we fixed *lpo2, upo2, lpco2*, and *upco2* to be 1.9, 14.8, 1.7 and 16.9, respectively, we only had to consider 5 other parameters in an ArtiDetect instance. With the setting of a possible value set for each of the 5 parameters, we defined a huge 5-dimensional Cartesian space $DPO2 \times DPCO2 \times CPCO2 \times CPO2 \times T$. Each element in the space, together with the fixed *lpo2, upo2, lpco2*, and *upco2,* determined an ArtiDetect instance. For example, it is easy to check that (1.01, 0.33, 0.198, 0.404, 6) is an element in that Cartesian space, and it determines an instance, i.e., ArtiDetect(1.9, 14.8, 1.7, 16.9, 1.01, 0.33, 0.198, 0.404, 6).

Table 3. Optimal ArtiDetect instances with different lengths of moving time window

t	*dpo2*	*dpco2*	*cpo2*	*cpco2*	sens(PO_2)/SD	spec(PO_2)/SD	sens(PCO_2)/SD	spec(PCO_2)/SD	*error*
3	0.86	0.18	0.258	0.108	0.889/0.079	0.957/0.038	0.969/0.048	0.948/0.035	0.236
3	0.86	0.19	0.258	0.114	0.889/0.079	0.957/0.038	0.969/0.048	0.948/0.035	0.236
3	0.86	0.21	0.258	0.105	0.881/0.078	0.960/0.033	0.963/0.055	0.960/0.029	0.236
3	0.86	0.22	0.258	0.110	0.881/0.078	0.960/0.033	0.963/0.055	0.960/0.029	0.236
3	0.86	0.23	0.258	0.115	0.881/0.078	0.960/0.033	0.963/0.055	0.960/0.029	0.236
3	0.91	0.18	0.273	0.108	0.889/0.078	0.960/0.034	0.966/0.049	0.950/0.034	0.236
3	0.91	0.19	0.273	0.114	0.889/0.078	0.960/0.034	0.966/0.049	0.950/0.034	0.236
3	0.92	0.18	0.273	0.108	0.889/0.078	0.960/0.033	0.966/0.049	0.950/0.034	0.236
3	0.92	0.19	0.273	0.114	0.889/0.078	0.960/0.033	0.966/0.049	0.950/0.034	0.236
3	1.01	0.18	0.303	0.108	0.882/0.074	0.966/0.030	0.966/0.049	0.951/0.033	0.236
3	1.01	0.19	0.303	0.114	0.882/0.074	0.966/0.030	0.966/0.049	0.951/0.033	0.236
3	1.02	0.18	0.306	0.108	0.880/0.075	0.967/0.029	0.966/0.049	0.951/0.032	0.236
3	1.02	0.19	0.306	0.114	0.880/0.075	0.967/0.029	0.966/0.049	0.951/0.032	0.236
3	1.03	0.18	0.309	0.108	0.879/0.074	0.968/0.029	0.966/0.049	0.951/0.032	0.236
3	1.03	0.19	0.309	0.114	0.879/0.074	0.968/0.029	0.966/0.049	0.951/0.032	0.236
3	1.04	0.18	0.312	0.108	0.879/0.074	0.968/0.029	0.966/0.049	0.951/0.032	0.236
3	1.04	0.19	0.312	0.114	0.879/0.074	0.968/0.029	0.966/0.049	0.951/0.032	0.236
4	1.04	0.24	0.312	0.144	0.913/0.060	0.959/0.033	0.972/0.041	0.951/0.036	0.204
5	1.06	0.27	0.318	0.162	0.930/0.049	0.950/0.039	0.975/0.044	0.944/0.038	0.200
5	1.06	0.27	0.318	0.189	0.930/0.049	0.950/0.039	0.971/0.048	0.948/0.035	0.200
6	1.01	0.33	0.404	0.198	0.950/0.045	0.942/0.045	0.972/0.036	0.941/0.042	0.195
7	1.08	0.37	0.324	0.185	0.949/0.036	0.935/0.048	0.974/0.031	0.930/0.048	0.212

Searching for optimal ArtiDetect instances

We used all 10 training infants as a training set to obtain the optimal ArtiDetect instances. After running for nearly 29 hours on an Ultra 5 Sun Microsystems workstation running Solaris 2.6, our search program found all the optimal ArtiDetect instances according to different lengths of the moving time window, as shown in Table 3.

The ArtiDetect instance with *t* of 6 in Table 3 was of particular interest. It was the best instance over all different lengths of the moving time window, because its error rate, i.e., 0.195, was the smallest. That is, if we set *t* to be 6, *dpo2* 1.01, *dpco2* 0.33, *cpo2* 0.404, and *cpco2* 0.198, and used the fixed settings for the other parameters, we obtained the optimal ArtiDetect instance, denoted by ArtiDetector, whose sens(PO_2) is 95.0% (SD = 4.5%), spec(PO_2) 94.2% (SD = 4.5%), sens(PCO_2) 97.2% (SD = 3.6%), and spec(PCO_2) 94.1 (SD = 4.2%). It was also interesting to note that, over all 10 training infants, ArtiDetector hit 97.0% and 98.0% of artifactual episodes in the PO_2 and PCO_2 channels, respectively. All the missed artifactual episodes consisted of only one or two value points. Figure 2 gives all the artifactual episodes of the chosen 10 infants identified by the expert and the ArtiDetector.

Note that Infant 8 had the most artifacts in both of the PO_2 and PCO_2 channels (164 and 124, respectively), but had relatively fewer artifactual episodes. For this particular infant, the sensitivity and specificity of ArtiDectector for detecting PO_2 artifacts was 93.3% and 95.1%, respectively; and the sensitivity and specificity for detecting PCO_2 artifacts 95.3% and 89.0%, respectively.

Finally, we want to answer two interesting yet difficult questions. The first question is that: Are the 10 infants sufficient for discovering optimal ArtiDetect instances? This question is hard to answer, because so many parameters are involved and extending our search space to progressively larger numbers of training infants would likely involve large multiples of the 29 hours of computer time of our original search. To answer that question, however, we randomly selected 8 infants out of the 10 infants, and used them as a second training dataset.* Our search system searched the defined search space for optimal ArtiDetect instances with this smaller training dataset. It ran for about 23 hours, and discovered all the best ArtiDetect instances based on different *t*'s, as shown in Table 4.

From Table 4, two conclusions can be made. First, the ArtiDetect instance with *t* of 6 in Table 4, denoted by ArtiDetector', is the best ArtiDetect instance over all the lengths of moving time window. For PCO_2 arti-

* The selected 8 infants are the 10 infants except the 4th and 8th infants in Table 2.

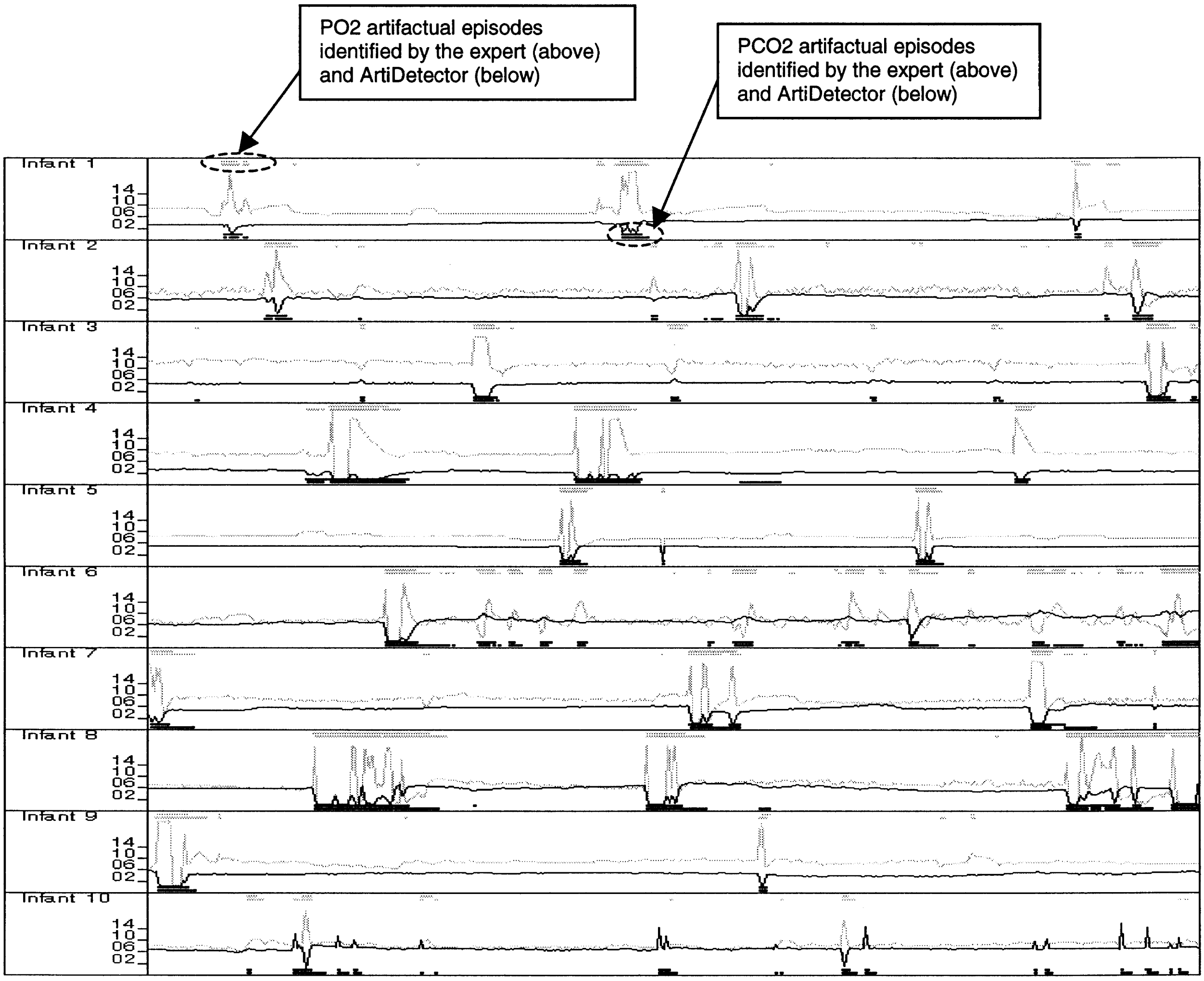

Fig. 2. All the Artifactual episodes in the 10-hour data segments of the 10 infants identified by the expert and ArtiDetector (The light and dark waves are PO_2 and PCO_2 data streams, respectively).

Table 4. A test on the size of training cases

t	*dpo2*	*dpco2*	*cpo2*	*cpco2*	sens(PO_2)/SD	spec(PO_2)/SD	sens(PCO_2)/SD	spec(PCO_2)/SD	*error*
3	0.91	0.18	0.273	0.108	0.885/0.080	0.955/0.036	0.961/0.054	0.950/0.039	0.249
3	0.91	0.19	0.273	0.114	0.885/0.080	0.955/0.036	0.961/0.054	0.950/0.039	0.249
3	0.92	0.18	0.276	0.108	0.885/0.080	0.956/0.036	0.961/0.054	0.950/0.039	0.249
3	0.92	0.19	0.276	0.114	0.885/0.080	0.956/0.036	0.961/0.054	0.950/0.039	0.249
4	0.94	0.24	0.283	0.144	0.920/0.061	0.947/0.040	0.973/0.045	0.949/0.041	0.211
4	1.02	0.24	0.306	0.144	0.912/0.056	0.953/0.036	0.973/0.045	0.951/0.040	0.211
4	1.03	0.24	0.309	0.144	0.912/0.056	0.953/0.036	0.973/0.045	0.951/0.040	0.211
4	1.04	0.24	0.312	0.144	0.912/0.056	0.954/0.035	0.973/0.045	0.951/0.040	0.211
5	1.06	0.27	0.318	0.189	0.930/0.042	0.945/0.042	0.968/0.053	0.948/0.039	0.208
6	1.01	0.33	0.303	0.196	0.955/0.034	0.931/0.048	0.971/0.040	0.941/0.048	0.202
7	1.08	0.37	0.324	0.185	0.950/0.029	0.927/0.051	0.974/0.033	0.928/0.054	0.220

Table 5. Optimal ArtiDetect instances without artifactual correlation

t	*dpo2*	*dpco2*	sens(PO_2)/SD	spec(PO_2)/SD	sens(PCO_2)/SD	spec(PCO_2)/SD	*error*
3	0.50	0.16	0.906/0.075	0.895/0.113	0.953/0.058	0.952/0.029	0.295
3	0.50	0.17	0.906/0.075	0.895/0.113	0.953/0.058	0.952/0.029	0.295
3	0.50	0.20	0.906/0.075	0.895/0.113	0.946/0.067	0.959/0.026	0.295
4	0.61	0.22	0.911/0.070	0.907/0.096	0.959/0.057	0.955/0.029	0.268
4	0.62	0.22	0.905/0.071	0.912/0.090	0.959/0.057	0.955/0.029	0.268
5	0.67	0.21	0.930/0.055	0.908/0.091	0.970/0.041	0.941/0.037	0.251
5	0.67	0.22	0.930/0.055	0.908/0.091	0.967/0.047	0.944/0.036	0.251
5	0.69	0.21	0.926/0.053	0.912/0.088	0.970/0.041	0.941/0.037	0.251
5	0.69	0.22	0.926/0.053	0.912/0.088	0.967/0.047	0.944/0.036	0.251
6	0.76	0.20	0.935/0.066	0.912/0.087	0.978/0.036	0.922/0.050	0.254
7	0.92	0.19	0.928/0.073	0.927/0.068	0.980/0.028	0.898/0.064	0.267

facts, ArtiDetector' and ArtiDetector have exactly the same sensitivity and specificity but ArtiDetector' has high sensitivity (95.5%) and specificity (93.1%) for detecting PO_2 artifacts. Second, in both of the best instances, *t, dpo2, dpco2*, and *cpco2* are exactly the same, but the *cpo2* values (0.404 *vs.* 0.303) are slightly different. The result provides some reassurance that the 10 training infants, with 10×600 records in total, should be sufficient enough to discover optimal ArtiDetect instances.

Table 5 examines whether artifactual correlation really matters in artifact detection. We first excluded the correlation-based detectors for the PO_2 and PCO_2 channels from consideration. By using the complete training set of 10 infants and an exhaustive search in the search space, we found the optimal ArtiDetect instances based on different *t*'s. According to the optimality criterion, each ArtiDetect instance with the least *error* value (i.e., 0.251) was our best ArtiDetect instance. Comparison to the results with the correlation detection showed that this optimal ArtiDetect instance was inferior to ArtiDetector where artifactual correlation was considerd. In addition, by comparing Table 3 and Table 5 we find that the *error* values of the ArtiDetect instances in Table 5 were consistently higher given a particular time window (i.e., with a fixed value for *t*). This suggests that the artifactual correlation between the PO_2 and PCO_2 channels play an important role in artifact detection.

DISCUSSION

Artifacts in time series monitoring data need to be identified and processed before meaningful conclusions can be made from the data. However, identifying such artifacts may be difficult. In our experience, it would be unusual for clinical notes to be complete enough to allow artifact detection by cross checking from the human-entered documentation. Thus, automatic identification is a necessary requirement for retrospective data analysis. It is also likely to be the only viable method for artifact eradication in real-time for pattern recognition of non-artifactual clinical events.

In this work, one expert served as the "gold standard". It is accepted that the gold standard may be "impure", but this has to be a first exploratory phase in the development of any automated artifact detectors. When multiple experts are available, the gold standard could be more reliable, and hence we could discover a more accurate artifact detector. However, as noted by Cunningham et al. [4], experts may not always agree on what artifacts are in a retrospective analysis, and thus there may be always some controversy in artifact detection.

ArtiDetector may fail in rare artifactual situations where PO_2 and PCO_2 are within the limits, and at the same time they are quite steady with a moving window. In this case, all the component artifact detectors in ArtiDetector may miss those situations. However, it should be pointed out that those situations are very likely to be preceded by *sudden* changes in PO_2 and PCO_2, which is detectable by ArtiDetector. For example, when the PO_2/PCO_2 probe is off an infant's body, PO_2 suddenly rises to about 20 kPa, and PCO_2 drops close to zero – these are the levels of oxygen and carbon dioxide in the air. The situation can certainly be detected by ArtiDetector.

By randomly selecting the training infants for experiments, we ignored some other potentially important features, e.g., weight, sex, gestational age and postnatal age. It is likely therefore that the discovered optimal artifact detector ArtiDetector is quite generally applicable. If on the other hand one considers a special group of training infants (e.g., infants whose gestational age is less than 26 weeks), we might discover an optimal artifact detector which reveals special artifactual behaviour in that group of infants.

We previously reported a simple and automatic method for detecting artifact in a single data stream that required no domain knowledge [3], and here we considered dual data streams. This is logical in this instance because the probe measuring PO_2 and PCO_2 is a combined probe, artifacts would often be related to the probe itself and would be reflected in both data streams. This gives the rationale for the correlation-based component of the artifact detector. We have only considered artifactual correlation among the PO_2 and PCO_2 channels, but use of some other channels that artifactually correlate with PO_2 and PCO_2 might improve the artifact detection in PO_2 and PCO_2 data streams. This is one of our future research goals.

In conclusion, our artifact detection method detected most PO_2 and PCO_2 in the 10 infants randomly selected from our clinical database. The method is simple and makes little use of domain knowledge. We believe that the method is easily extensible to detect artifacts in other channels (e.g., the heart rate and blood pressure channels) when a proper gold standard is established for artifacts in those channels.

This work is supported in part by a DARPA grant # F30602-97-1-0193. We would like to thank Dr Atul Butte, Dr Daniel Nigrin, and Professor Peter Szolovits for their valuable comments on this work.

GLOSSARY

PCO_2	transcutaneous partial pressure of oxygen
PO_2	transcutaneous partial pressure of carbon dioxide

REFERENCES

1. Meredith C, Edworthy J. Are there too many alarms in the intensive care unit? An overview of the problems. Int J Adv Nurs 1995; 21: 15–20
2. Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. Crit Care Med 1997; 25: 614–619
3. Cao CG, McIntosh N. Empirical study on artifact detection in monitoring data. In Proceedings of the 1998 Annual Symposium of American Medical Informatics Association, 1998; 983
4. Cunningham S, Symon AG, McIntosh N. The practical management of artifact in computerised physiological Data. Int J Clin Monit Comput 1994; 11: 211–216
5. Sittig DF, Factor M. Physiological trend detection and artifact rejection: A parallel implementation of a multistate Kalman filtering algorithm. Int J Comput Meth Prog in Bio 1990; 31: 1–10
6. Rampil IJ. Intelligent detection of artifact. In: Gravenstein JS, Newbower RS, Ream AK et al., eds. The Automated Anesthesia Record and Alarm System. Boston: Butterworths, 1987: 175–190
7. Young WH, Gardner RM, East TD, Turner K. Computerized ventilator data selection: Artifact rejection and data reduction. Int J Clin Monit Comput 1994; 14: 165–176
8. Rheineck-Leyssius AT, Kalkman CJ. Influence of pulse oximeter settings on the frequency of alarms and detection of hypoxemia: Theoretical effects of artifact rejection, alarm delay, averaging, median filtering or a lower setting of the alarm limit. Int J Clin Monit Comput 1998; 14: 151–156
9. Poets CF, Stebbens VA. Detection of movement artifact in recorded pulse oximeter saturation. Eur J Pediatr 1997; 156: 808–811