Reduction of false arterial blood pressure alarms using signal quality assessment and relationships between the electrocardiogram and arterial blood pressure

W. Zong G. B. Moody R. G. Mark

Harvard-MIT Division of Health Sciences & Technology, Massachusetts Institute of Technology, Cambridge, USA

Abstract—*The paper presents an algorithm for reducing false alarms related to changes in arterial blood pressure (ABP) in intensive care unit (ICU) monitoring. The algorithm assesses the ABP signal quality, analyses the relationship between the electrocardiogram and ABP using a fuzzy logic approach and post-processes (accepts or rejects) ABP alarms produced by a commercial monitor. The algorithm was developed and evaluated using unrelated sets of data from the MIMIC database. By rejecting 98.2% (159 of 162) of the false ABP alarms produced by the monitor using the test set of data, the algorithm was able to reduce the false ABP alarm rate from 26.8% to 0.5% of ABP alarms, while accepting 99.8% (441 of 442) of true ABP alarms. The results show that the algorithm is effective and practical, and its use in future patient monitoring systems is feasible.*

Keywords—*Patient monitoring, False alarms, Intensive care unit, Arterial blood pressure, ECG, Fuzzy logic*

Med. Biol. Eng. Comput., 2004, **42**, 698–706

1 Introduction

INTENSIVE CARE unit (ICU) monitors generate a high incidence of false alarms, which becomes an annoying problem (WATT *et al.*, 1993; LAWLESS, 1994; TSIEN and FACKLER, 1997). Clinicians sometimes solve this problem by simply disabling the alarms altogether. A better solution, however, would be to reduce the incidence of false alarms, without missing true alarm events.

ICU monitors most often produce false alarms when a physiological signal is corrupted by artifacts. Although most such false alarms can be easily identified by looking at the signal quality and by referencing other related signals, ICU monitors do not typically have sophisticated signal quality analysis, and they generally do not take advantage of the known relationships between signals of different modalities.

Fig. 1 shows examples of arterial blood pressure (ABP) artifacts found in the MIMIC Database (MOODY and MARK, 1996). It should be noted that some artifacts, such as those in Fig. 1*d*, can be very similar in appearance to real physiological changes.

Previous efforts have been made to reduce ICU false alarms based on analysis of single or multichannel measurements available from commercial monitors (CREW *et al.*, 1991; MAKIVIRTA *et al.*, 1991; FELDMAN *et al.*, 1997; RHEINECK-LEYSSIUS and KALKMAN, 1998; TSIEN and KOHANE, 1998; CAO *et al.*, 1999; TSIEN *et al.*, 2000). Although these studies did not directly address the quality of the signals from which the measurements were derived, their results encourage further research.

This study presents an algorithm for reducing false ABP alarms by assessing the signal quality of the ABP waveform and by fusing information from simultaneous electrocardiogram (ECG) and ABP signals (ZONG *et al.*, 1999). The process employed a fuzzy logic analysis approach. We used separate subsets of records from the MIMIC database for development and for evaluation. Our results suggest that this algorithm is effective and practical and shows promise for use in future patient monitoring systems.

2 Materials and methods

2.1 *Database*

We used 25 multi-parameter records of ICU patients from the MIMIC database, which is freely available from PhysioNet (http://www.physionet.org) (GOLDBERGER *et al.*, 2000), for the algorithm development set, and 28 different records from the same database for the algorithm evaluation test set. The development set consisted of a total of 825 h of data taken from 20 patients (ten male, ten female, age 52–92 years). The record

Correspondence should be addressed to Dr Wei Zong; email: wzong@mit.edu

Paper received 17 November 2003 and in final form 28 April 2004

MBEC online number: 20043912

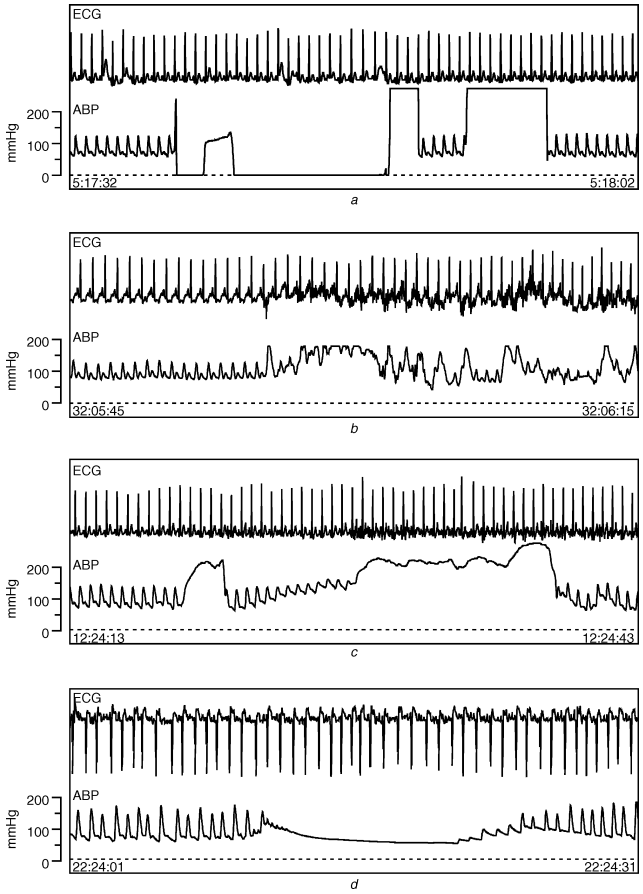

Fig. 1 *Examples of artifacts in ABP signals: (a) transducer flushing, (b) motion, (c) probably movement induced, (d) proximal BP cuff inflation. 30 per trace*

duration averaged 33 h (ranging from 8.7 to 62.7 h in length). The test set records were from 26 patients (16 male, ten female, age 21–92 years) and totalled 1065 h in duration. The record duration averaged 38 h (ranging from 10.6 to 58.5 h).

Patients involved in this study had a wide range of clinical problems, including sepsis, respiratory failure, congestive heart failure/pulmonary oedema, haemorrhage, brain injury, myocardial infarction, cardiogenic shock and post-operative care for cardiac surgery. All patient records contained at least multi-lead ECGs and radial ABP, which were used in this study. Most records contained additional physiological signals, such as pulmonary arterial pressure, pulse oximetry (plethysmograph), impedance-based respiration etc. The ECG signals were digitised at 500 Hz, and other signals were digitised at 125 Hz, with 12-bit resolution. The records also contained extensive clinical data, including patient history, laboratory results, medication, fluid balance and clinical observations.

In addition to the physiological signals and the clinical data, the MIMIC database includes alarm annotations produced by the bedside monitors. When an alarm condition (a measurement crossing a threshold) is first detected, alarm annotations are generated at intervals of 1.024 s, and they continue until either the measurement returns to an acceptable range or the ICU staff intervenes to silence the alarm. In this study, alarm annotations referring to the same alarm condition were treated as a single alarm event. A new alarm condition would be defined only after an interval of 15 s without an alarm. A few alarm annotations that were associated with saturated digital signals were removed, because correct ABP measurements were not available (from the saturated signal). Each machine-annotated alarm condition was carefully examined by the authors, without reference to our algorithm's classification of the event, and judged to be either true or false.

The criteria for manually annotating ABP alarms were as follows: For each monitor-generated ABP alarm, at least 30 s (up to 5 min) of ABP and ECG waveforms prior to the alarm were scrutinised. The ABP measurements from the ABP waveform in this region were also manually checked by means of a graphic signal viewing/measuring tool called WAVE (MOODY and MARK, 1991), which is freely available from PhysioNet. If the ABP signal was corrupted by artifacts (as identified by human experts) resulting from events such as catheter flush, patient movement or a clot-blocked transducer, the alarm was annotated as a false alarm. If the ABP signal was clean, and there was no sign of a transducer-caused problem, the manually checked measurements from the waveform matched the values obtained from the monitor for the alarm, and the changes in the ABP signal could be understood by reference to the ECG, the alarm was considered a true alarm. When the ABP signal had changes that could be either physiological or artifact-related, the ECG reference was very important in verifying the cause of the ABP changes; see Fig. 1*d* and Fig. 9*f*.

The development set contained a total of 445 ABP alarm events, of which 319 were true positives and 126 were false positives (28.3%). The test set contained 604 ABP alarm events, of which 442 were true positives and 162 were false (26.8%) positives.

2.2 *Methodology overview*

Our approach was based on beat-by-beat ABP signal quality analysis with incorporation of ECG–ABP relationships. The structure of the algorithm is shown in Fig. 2.

First, ABP signal quality was assessed on a beat-by-beat basis, yielding a signal quality index SQI_i, that had a value between 0 and 1. SQI_i was derived from and was associated with each beat. Secondly, if QRS information from ECG data was available, SQI_i could be modified, based on analysis of the ECG rhythm and ECG–ABP delay time (the time delay between the QRS and the following ABP pulse). Thirdly, only ABP measurements (systolic, diastolic and mean blood pressure) derived from

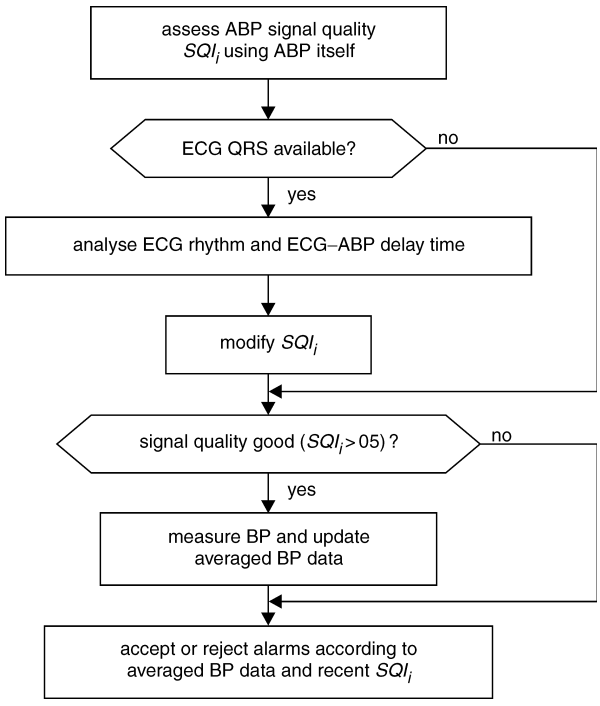

Fig. 2 *Overview of algorithm*

high-quality signals (those signals with high SQI_i value) were used for updating the short-term averaged blood pressure (BP) values. Finally, acceptance or rejection of each alarm was based on the short-term averaged BP values and recent SQI_i values preceding each alarm.

2.3 *Signal quality assessment of the ABP*

Signal quality assessment of the ABP was performed through beat-by-beat fuzzy testing of ABP waveform features. This process consisted of ABP pulse detection, waveform feature extraction, waveform feature fuzzy representation and fuzzy reasoning to produce the signal quality index.

2.3.1 *ABP pulse detection algorithm:* An effective ABP pulse onset detection algorithm was developed based on the regional slope feature of the low-pass filtered ABP signal (ZONG *et al.*, 2003). The main ascending portion (from the onset to the peak) of the ABP pulse possesses the maximum average positive slope. A slope sum function (SSF) was defined as shown in (1) to enhance the ascending portion of the ABP pulse and to suppress the remainder of the ABP signal, to simplify the process of ABP pulse detection.

$$SSF(k) = \sum_{i=k-w}^{k} \Delta y_i \quad \Delta y_i = \begin{cases} \Delta x_i & \text{if } \Delta x_i > 0 \\ 0 & \text{if } \Delta x_i \leq 0 \end{cases} \tag{1}$$

In (1), k is the current sample number, w is a window that is approximately equal to the duration of the ascending portion of a typical ABP pulse (we chose $w = 16$, or 128 ms at the 125 Hz sampling rate), $\Delta x_i = x_i - x_{i-1}$, and x_i is the ith ABP sample.

The relationship between the original ABP and the transformed signal *SSF* is shown in Fig. 3. The onset of the SSF pulse corresponds with the onset of the ABP pulse, as the *SSF* signal rises only when the ABP signal rises or when noise in the signal is not suppressed by the low-pass filter. The ABP pulse can be detected through observation of the SSF pulse.

The decision rule for detecting the SSF pulse onset consisted of two procedures: adaptive thresholding of the SSF signal to detect SSF pulses of appropriate amplitude, and local searching around the detection point to confirm the detection and to identify the onset of the pulse. During the thresholding step, a threshold base value was established and was initialised at three times the mean SSF signal (averaged over the first 1000 samples of recording). The threshold base value was adaptively updated by the maximum SSF value for each SSF pulse detected. The actual threshold was taken to be 60% of the threshold base value. When the SSF signal crossed this threshold, the algorithm searched for the minimum and the maximum values in a 150 ms window preceding and succeeding the threshold-crossing point, respectively. The pulse detection was accepted only if the difference between the maximum and minimum exceeded a predefined, empirical value. When the pulse was accepted, the algorithm searched backward in time from the threshold-crossing point for the onset of the SSF pulse. The onset point was determined when the SSF signal dropped to the minimum +1.0% of the maximum SSF value. The calculated ABP pulse onset was adjusted by 16 ms, or two samples, to compensate for the low-pass filter's phase shift. Finally, to avoid double detection of the same pulse, a 250 ms eye-closing (refractory) period was applied, during which no new pulse detection was initiated.

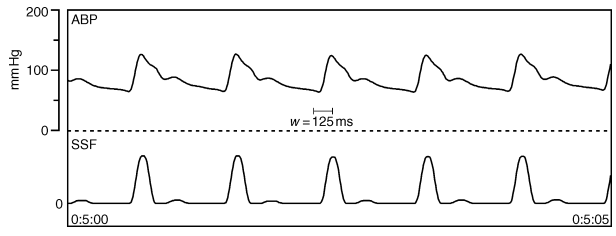

Fig. 3 *Typical ABP waveform from MIMIC database and corresponding slope sum function (SSF)*

2.3.2 *ABP waveform feature extraction:* ABP waveform feature extraction was performed on a beat-by-beat basis. The waveform features used in this study were systolic blood pressure (SBP), diastolic blood pressure (DBP), mean blood pressure (MBP), maximum positive pressure slope (MPPS), maximum negative pressure slope (MNPS), maximum up-slope duration (MUSD) (which was the maximum duration that the ABP signal continued rising), maximum duration above threshold (MDAT) (which was the maximum duration that the ABP signal stayed above a threshold), pulse-to-pulse interval (PP), pulse blood pressure (PBP) (which was the difference between SBP and DBP in a beat) and ECG-ABP delay time (DT) (which was the interval between the QRS onset in the ECG and the onset of the following ABP pulse; this feature could be obtained only when ECG was available).

At the beginning of the record there was a 20 s learning period. During this learning period, each ABP pulse was detected and, in each beat cycle (from the previous pulse onset to the current one), the waveform features listed above were extracted and averaged. The averaged features established in the learning period were the initial base feature set.

After the learning period, the algorithm worked as shown in Fig. 4. For beat cycle i, a 2 s detection time window was defined from the previous window endpoint. If the previous window ended at a detected pulse (i.e. if the flag was set to 1), the detection window started at t_0 plus an eye-closing period (Δt, 250 ms) to jump over the previous detected pulse. Otherwise, the window began at t_0, as there was no pulse detected in the previous window. If an ABP pulse was detected within the window (between t_0 and t_1), the end of the window t_1 was reset to the time of the onset of the detected pulse. Otherwise, the window ended as previously set (i.e. 2 s from t_0). The flag was set to 1 or 0, according to whether a pulse was detected or not. The

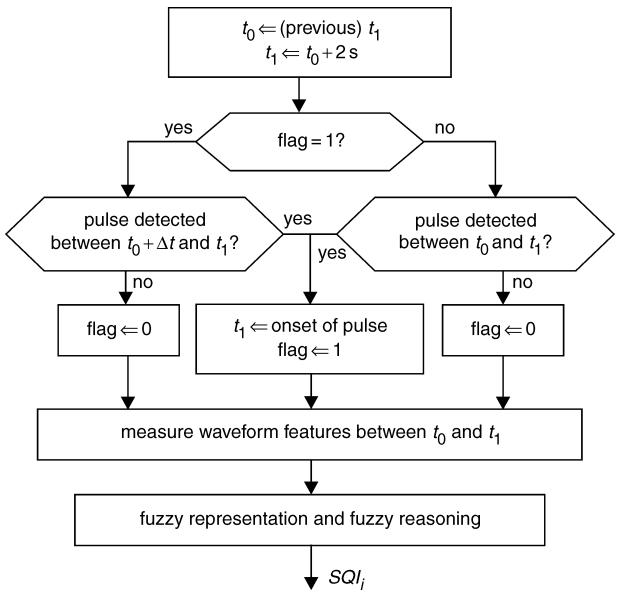

Fig. 4 *Assessment of ABP signal quality using ABP itself*

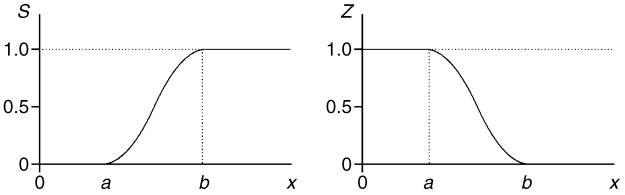

Fig. 5 *Shapes of S-function and Z-function*

features of the waveform within the window, from t_0 to t_1, were extracted and kept for later analysis.

2.3.3 *Fuzzy representation and reasoning:* Based on fuzzy set theory and applications (ZADEH, 1965, 1983; PEDRYCZ and GOMIDE, 1998; ZONG and JIANG, 1998) and knowledge of the ABP waveform morphology (O'ROURKE *et al.*, 1992; BERNE and LEVY, 1997), a group of linguistic variables was defined to describe the local waveform characteristics: '*ABP_amplitude_too_high*', '*ABP_slope_normal*', '*ABP_keeps_rising_too_long*' etc. Two standard fuzzy set membership functions, the *S* function (PEDRYCZ and GOMIDE, 1998) and the *Z* function, defined in (2) and (3) and shown in Fig. 5, were used to define these linguistic variables (see Table 1).

During the processing cycle, all the linguistic variables were calculated based on the extracted waveform features and the reference values in the base feature set, which evolved as the signal changed (see Section 2.5 for details).

$$S(x; a, b) = \begin{cases} 0 & x \leqslant a \\ 2\left(\dfrac{x-a}{b-a}\right)^2 & a < x \leqslant \dfrac{a+b}{2} \\ 1 - 2\left(\dfrac{x-b}{b-a}\right)^2 & \dfrac{a+b}{2} < x \leqslant b \\ 1 & b < x \end{cases} \quad (2)$$

$$Z(x; a, b) = 1 - S(x; a, b) \quad (3)$$

From these linguistic variables, three composite variables were defined using fuzzy conditional statements (FCSs): *ABP_amplitude_normal* (*AN*), *ABP_slope_normal* (*SN*), and *ABP_with_blocked_transducer* (*WBT*). These statements are as follows:

IF [**not** '*ABP_amplitude_too_large*' (*ATL*)] **and**
[**not** '*ABP_amplitude_too_small*' (*ATS*)]
THEN '*ABP_amplitude_normal*' (*AN*)

$$\mu_{AN} = (1 - \mu_{ATL}) \wedge (1 - \mu_{ATS}) \quad (4)$$

where operator $\wedge$ is the standard fuzzy intersection (ZADEH, 1965); and $\mu_A(x) \wedge \mu_B(x) = min[\mu_A(x), \mu_B(x)]$, for all x in the appropriate domain.

IF [**not** '*ABP_slope_too_large*' (*STL*)] **and**
[**not** '*ABP_slope_too_small*' (*STS*)]
THEN '*ABP_slope_normal*' (*SN*)

$$\mu_{SN} = (1 - \mu_{STL}) \wedge (1 - \mu_{STS}) \quad (5)$$

IF ['*ABP_pulse_pressure_decrease*' (*PPD*)] **and**
['*ABP_diastolic_pressure_increase*' (*DBPI*)] **and**
[**not** '*premature_ABP_pulse*' (*PrP*)]
THEN '*ABP_with_blocked_transducer*' (*WBT*)

$$\mu_{WBT} = \mu_{PPD} \wedge \mu_{DBPI} \wedge (1 - \mu_{PrP}) \quad (6)$$

Finally, a conclusive linguistic variable *signal_quality_good* (*SQG*) was defined to describe the signal quality in the localised period. The signal quality index, SQI_i, was assigned as the certainty degree of *SQG*, as seen in (7).

IF ['*ABP_amplitude_normal*' (*AN*)] **and**
['*ABP_slope_normal*' (*SN*)] **and**
[**not** '*ABP_keeps_rising_too_long*' (*KRTL*)] **and**
[**not** '*ABP_stays_high_too_long*' (*SHTL*)] **and**
[**not** '*ABP_with_blocked_transducer*' (*WBT*)]
THEN '*signal_quality_good*' (*SQG*)

$$SQI_i = \mu_{SQG} = \mu_{AN} \wedge \mu_{SN} \wedge (1 - \mu_{KRTL}) \wedge (1 - \mu_{SHTL}) \wedge (1 - \mu_{WBT}) \quad (7)$$

The base feature set, established during the initial learning period, evolved using weighted averaging as the ABP waveforms changed. Only those ABP episodes with good signal quality ($SQI_i > 0.5$) were counted into the weighted averaging process.

2.4 *Use of ECG–ABP relationships*

The ECG and ABP signals are closely related in terms of rhythm and timing. When an ECG signal was available, the SQI_i, as described above, could be modified, based on analysis of the ECG–ABP relationships. Fig. 6 summarises this procedure.

The analysis of the ECG–ABP relationship began after the SQI_i was derived from the current ABP episode. If the current ABP episode contained a detected ABP pulse (flag = 1), then the preceding QRS complex was checked. If the preceding QRS was not premature, and the ECG–ABP delay time fell within the expected range, indicating that the ABP pulse was associated with a real beat, or if the beat was premature (in which case the algorithm did not attempt to predict the timing of the ABP pulse), then SQI_i was not modified. Otherwise, the pulse could be an artifact, and SQI_i was set to zero. If no ABP pulse was detected in

Table 1 Linguistic variables and their definitions

Variable name	Description and definition	Parameter explanation
ATL	*ABP_amplitude_too_large:* $\mu_{ATL} =$ S *(SBP − SBPa; 20, 60)*	*SBP*: systolic BP, mmHg; *SBPa*: systolic BP base
ATS	*ABP_amplitude_too_small:* $\mu_{ATS} = Z$ (*DBP*; 0, 20)	*DBP*: diastolic BP, mmHg
STL	*ABP_slope_too_large:* $\mu_{STL} = S$ (*MPPS*/*MPPSa*; 1, 3)	*MPPS*: maximum positive BP Slope; *MPPSa*: MPPS base
STS	*ABP_slope_too_small:* $\mu_{STS} = S$ (*MNPS*/*MNPSa*; 1, 3)	*MNPS*: maximum negative PB slope; *MNPSa*: MNPS base
KRTL	*ABP_keeps_rising_too_long:* $\mu_{KRTL} = S$ (*MUSD*; 200, 500)	*MUSD*: maximum up-slope duration, ms
SHTL	*ABP_stays_high_too_long:* $\mu_{SHTL} = S$ (*MDAT*; 400, 800);	*MDAT*: maximum duration up threshold, ms
PPD	*ABP_pulse_pressure_decrease:* $\mu_{PPD} = Z$ (*PBP*/*PBPa*, 0.5, 0.9);	*PBP*: pulse blood pressure; *PBPa*: PBP base
DBPI	*ABP_diastolic*_pressure_*increase:* $\mu_{DBPI} = S$ *(DBP/DBPa; 0.8, 1.1);*	*DBP*: diastolic blood pressure: *DBPa*: DBP base
PrP	*Premature_ABP_pulse:* $\mu_{PrP} = Z$ (*PP*/*PPa*; 0.75, 0.95)	*PP*: pulse-pulse interval; *PPa*: PP base value

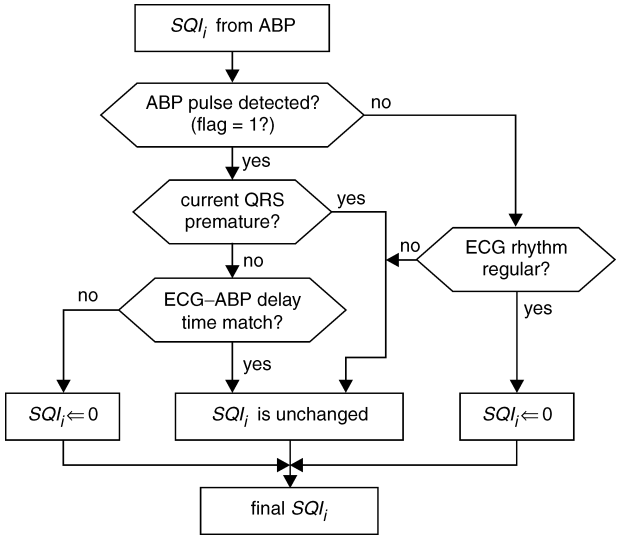

Fig. 6 *Procedure for ABP SQI_i modification by use of ECG–ABP relationship*

the current ABP episode, but the ECG rhythm was regular, this suggested the ABP non-pulse episode was due to artifact, and the SQI_i was set to zero. If the ECG rhythm was irregular, the SQI_i was accepted without modification, as such rhythms can be accompanied by the loss of ABP pulses.

In this study, a previously developed ECG beat detection and classification algorithm, Aristotle (MOODY and MARK, 1982), was employed to obtain QRS times of occurrence and types of QRS complex. According to the QRS occurrence time and QRS type obtained from Aristotle, the prematurity of the current QRS and the regularity of the current ECG rhythm can be determined.

To identify premature QRS complexes, a fuzzy variable '*premature_QRS* (*PrQ*)' was defined as

$$\mu_{PrQ} = Z(RR/RRa;\ 0.75,\ 0.95) \tag{8}$$

where *RR* was the current RR interval, and *RRa* was the recent short-term averaged RR interval. If $\mu_{PrQ} > 0.5$, the current beat was considered as a premature QRS.

The variable '*delay_time_match* (*DTM*)' was used to determine whether the QRS and the detected ABP pulse were a match. DTM was defined as follows:

$$\mu_{DTM} = S(DT/DTa;\ 0.4,\ 0.9) \wedge Z(DT/DTa;\ 1.1,\ 1.6) \tag{9}$$

where *DT* was the current QRS-ABP delay time, and *DTa* was the averaged QRS-ABP delay time established during the learning period.

The linguistic variable '*QRS_on_time* (*QOT*)' was defined as

$$\mu_{QOT} = S(RR/RRa;\ 0.75,\ 0.95) \wedge Z(RR/RRa;\ 1.05,\ 1.25) \tag{10}$$

If $\mu_{QOT} > 0.5$, the QRS was considered as on time. Aristotle provides QRS labels including normal (N), ventricular premature contraction (VPC), supra-ventricular premature contraction (SVPC) etc.

To determine if the ECG rhythm was regular, the 15 s period ending at the current time was considered. If, in this range, more than half of the QRS complexes were either on time or labelled as normal, the ECG rhythm was considered regular.

After the ECG–ABP relationship analysis procedure, the final SQI_i was acquired to determine the effective short-term ABP measurements and to update the base feature set, as discussed in the following Section.

Fig. 7 shows an example of the ABP signal quality index (SQI) derived from the ECG and ABP data from record 254 in

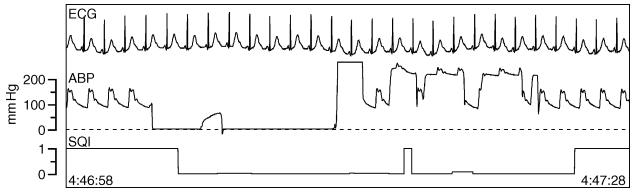

Fig. 7 *Example of ABP signal quality index (SQI) (30 s). SQI signal lags ABP signal by one beat interval*

the MIMIC database. Low values of SQI correspond to poor ABP signal quality (note that there is a one-beat delay from the ECG and ABP signals to the SQI signal).

2.5 *Measurements from good-quality signals*

After the final SQI_i was determined, the base features were updated with the current waveform feature values if the corresponding SQI_i was high enough (>0.5). Not all base features were updated. The features that needed updating were *SBPa*, *DBPa*, *MBPa*, *MPPSa*, *MNPSa* and *PPa*. The updating mechanism is shown in (11).

$$NNa = 0.875^{*}Na + 0.125^{*}N \tag{11}$$

where *Na* was the previous value in the base feature set, *N* was the feature value from the current waveform, and *NNa* was the updated base feature value.

The algorithm obtained the instantaneous ABP measurements (systolic, diastolic and mean blood pressures) for each beat or episode when it completed the SQI analysis. The algorithm also produced the short-term averaged ABP measurements derived from instantaneous measurements, with $SQI_i > 0.5$. Values associated with poor signal quality were not counted in the averaged measurements (see Fig. 8). The short-term averaging method is given by (11). For each beat, if $SQI_i > 0.5$, the current average value *NNa* was primarily based on the previous averaged value *Na*, with a small adjustment based on the current value *N*. If $SQI_i \leq 0.5$, the average value was not modified. The output of the algorithm included the instantaneous beat-by-beat measurements, SQI_i, and the short-term averaged ABP measurements. The *SQI* and the short-term averaged ABP measurements were used as a basis for the algorithm's decisions regarding the alarms.

When we compared the ABP measurements produced by the monitor and those derived from our algorithm (see the example in Fig. 8, with data from MIMIC record 212), we could see that most unexpected spikes in the monitor's measurements had been removed.

2.6 *Criteria for reducing false ABP alarms*

The patient monitor used for collecting MIMIC records produced ABP alarms based on some sort of short-term averaged systolic ABP measurements. There was a delay, of about 10 s, before the monitor issued an alarm annotation.

Our algorithm judged the ABP alarms produced by the monitor based on the SQI_i and the averaged ABP measurements. This judgment was based on the 15 s prior to the onset of the alarm condition annotated by the monitor. If all the SQI_i in this 15 s interval were good ($\geqslant 0.5$), suggesting that the signal quality in this region was not a problem, the alarm was judged as true. If there were four or more bad SQI_i (<0.5) in this interval, indicating that the signal quality was bad, the alarm was judged as false. If there were up to three bad SQI_i, then the systolic ABP measurement from the monitor and the averaged systolic ABP measurements from the algorithm were compared. If at least three of our algorithm's measurements

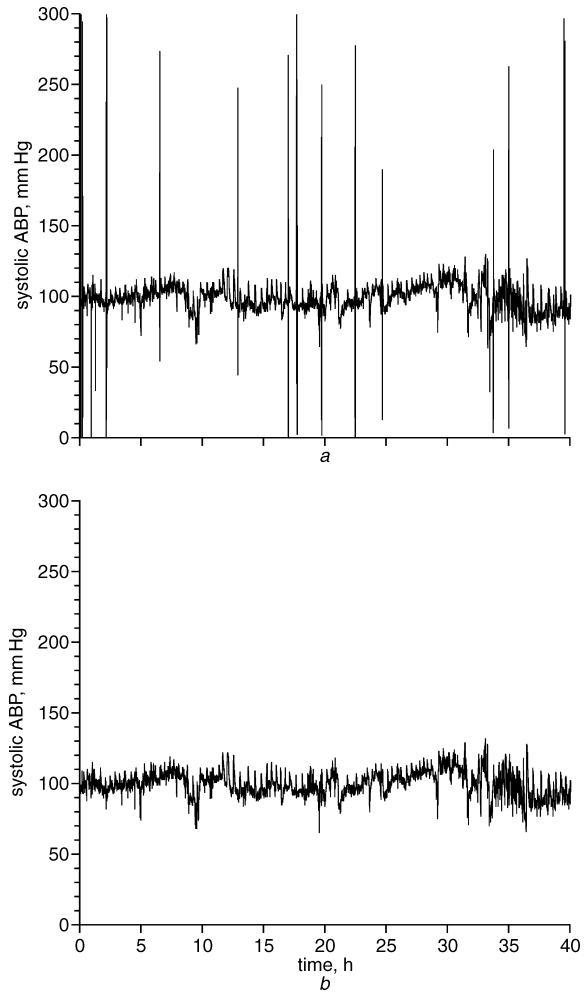

Fig. 8 *Systolic blood pressure against time: (a) SBP measurements produced by monitor; (b) SBP measurements from our algorithm*

were within 10 mmHg of the monitor's measurement, then the alarm was judged as true. This situation means that the signal quality in this episode was balanced between good and bad. The monitor that produced the alarm, based its measurements on the recent ABP signal. If the monitor's measurement was close enough to the averaged measurements derived from our algorithm, then the monitor's measurement was confirmed, and the alarm was accepted.

3 Results

As detailed in Table 4 in the Appendix, the development data set contained 25 records averaging 33 h in length. These ranged between 8.7 and 62.7 h (a total of 825 h). The monitors produced 445 ABP alarms, of which 126 (28.3%) were false based on our visual review. It was found that the algorithm rejected 117 of these 126 false alarms (92.9%) and only two true alarms (as summarised in Table 2). Thus the algorithm reduced the false alarm rate in the development data set from 28.3% to 2.8%, at a cost of rejecting 0.4% of 319 true alarms.

Table 2 Algorithm performance on the development data

Algorithm		true	false	total
Truth	true	317	2	319
	false	9	117	126

Sensitivity: 99.4%
Positive predictive accuracy: 97.2%

Table 3 Algorithm performance on the test data

Algorithm		true	false	total
Truth	true	441	1	442
	false	3	159	162

Sensitivity: 99.8%
Positive predictive accuracy: 99.3%

We repeated this experiment using the test data set, 28 records averaging 38 h in length, with a range of 10.6–58.5 h (1065 h in all). Table 5 in the Appendix shows the detailed evaluation results for the test data. The monitors produced 604 ABP alarms, of which 162 (26.8%) were false based on our visual review. The algorithm rejected 159 (98.2%) of these false alarms, reducing the false alarm rate from 28.6% to 0.4%, while rejecting only one (0.2%) of the 442 true alarms. Table 3 has a summary of the algorithm's performance.

Fig. 9 includes examples of false alarms rejected (Figs 9*a–c*), false alarm accepted (Fig. 9*d*), true alarms accepted (Figs 9*e* and *f*) and true alarms rejected (Figs 9*g–i*) from both the development and test data sets.

4 Discussion

The algorithm performed very well on the development set, rejecting almost all of the false positive alarms. Significantly, its performance on the test set was even better, supporting the hypothesis that the algorithm can be applied usefully to real-world data that have not been used for development. Any algorithm designed to reject false positives, however, can be expected to reject some true positives as well. Of particular concern in this study was the possibility that a clinically significant true alarm could be erroneously rejected. For this reason, we carefully examined the three cases in which true alarms were rejected by our algorithm (Figs 9*g–i*).

The two true alarms rejected by the algorithm in the development data set were distorted signal waveforms with real ABP changes; one was with hypertension, as shown in Fig. 9*g*, and the other was with hypotension, Fig. 9*h*. In Fig. 9*g*, there is transient hypertension accompanying patient movement. The alarm limit was 220 mmHg (for systolic ABP), and this alarm was annotated as a true alarm. The algorithm judged the ABP signal quality as not perfect (two beats with low SQI value), and the ABP measurements obtained by the algorithm did not meet the alarm limit at the time of the monitor's alarm; thus the algorithm marked the alarm as a false alarm. In the case of Fig. 9*h*, a hypotension alarm was present with the alarm limit at 85 mmHg (for systolic ABP). A catheter flush just prior to the alarm event caused the algorithm to deem the ABP signal quality low enough (four or more beats with low SQI value) to reject the alarm. Arguably, neither of these two cases was clearly a true alarm.

The single rejected true alarm in the test data set occurred when the systolic ABP suddenly increased following a long, slow attenuation of the ABP signal (see Fig. 9*i*). In this case, the algorithm adapted to the attenuated ABP signal and determined its signal quality to be good; when the signal suddenly increased to normal scale, the algorithm was not able to recognise the cause; it calculated a low SQI and thus rejected the alarm. This situation might be avoided by introducing an extra waveform base feature set and additional rules into the algorithm. The second feature set would keep the waveform features that are from the patient's most normal state. When the SQI derived from the first base feature set became low for a certain time, the second feature set would be

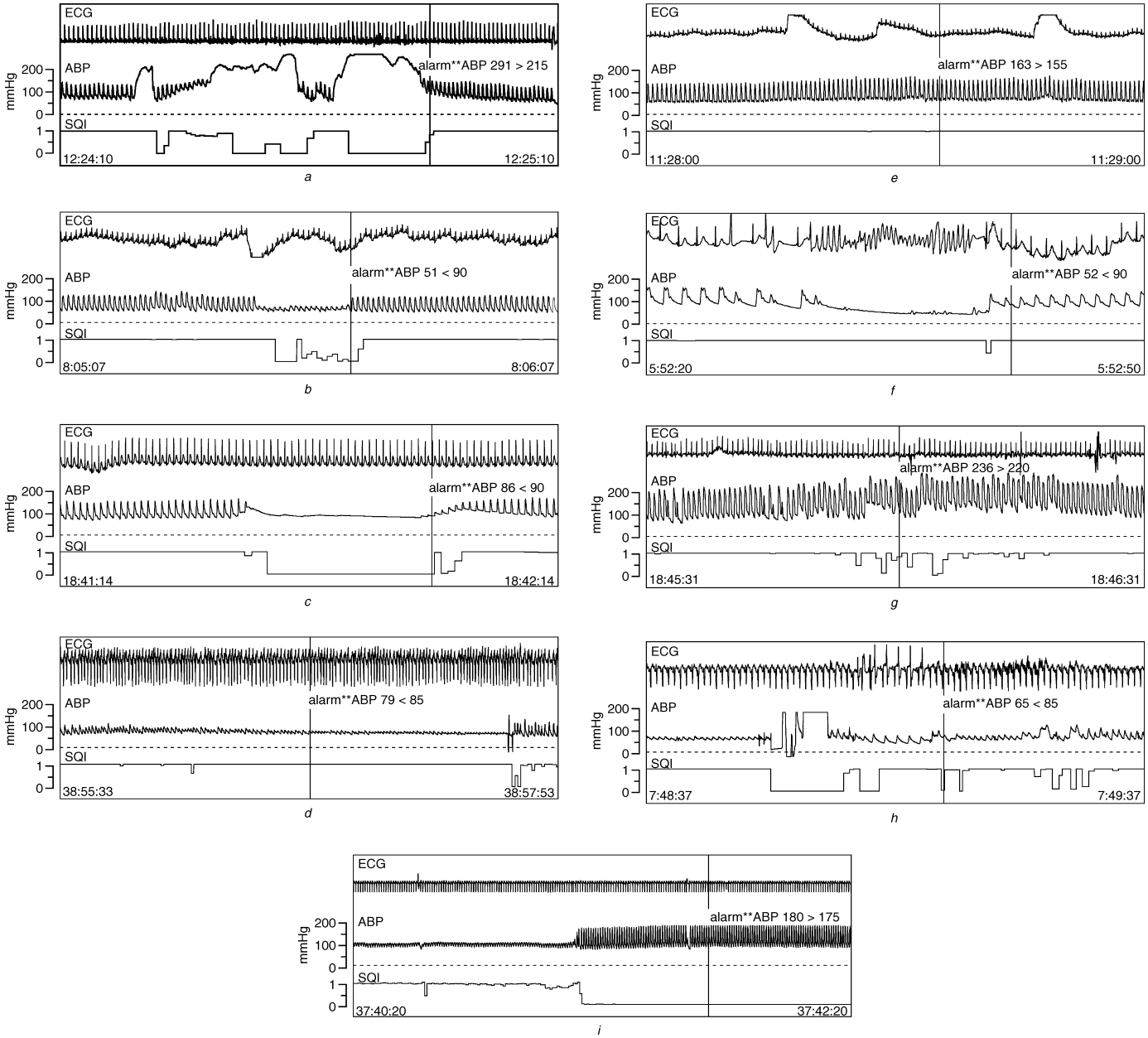

Fig. 9 *Examples of algorithm's judgment on ABP alarms: (a) false alarm caused by artifact (probably movement induced), rejected (four or more bad SQI_i), 1 min trace; (b) false alarm caused by obstructed catheter, rejected (four or more bad* SQI_i*), 1 min trace; (c) false alarm caused by cuff inflation, rejected (four or more bad SQI_i), 1 min trace; (d) false alarm due to clot-blocked transducer, accepted (SQI_i are all good within 15 s), 2 min trace; (e) true alarm related to ABP increase, accepted (SQI_i are all good), 1 min trace; (f) true alarm related to ventricular fibrillation, accepted (one bad SQI_i, BP measurements meet alarm limit), 30 s trace; (g) true alarm related to ABP increase and patient movement, rejected (two bad SQI_i, BP measurements do not meet alarm limit), 1 min trace; (h) true alarm right after catheter flush, rejected (four or more bad SQI_i), 1 min trace; (i) true alarm related to sudden increase in ABP signal, rejected (four or more bad SQI_i), 2 min trace*

used to verify the low calculated SQI of the current ABP signal. If the SQI from the second feature were good (i.e. >0.5), then the first obtained SQI would be modified, and the existing feature set would be updated from the current waveform features. In this way, the algorithm could avoid misbehaving when the ABP signal returns to a normal clean state from a damped situation, as in Fig. 9*i*.

The nine false alarms that the algorithm did not reject in the development set were due to ABP signals obtained while the transducer was blocked by blood clots. Such signals show the systolic ABP gradually dropping and the mean and diastolic ABP staying almost the same. The algorithm includes special fuzzy conditional statements, see (6) and (7), for such situations. As a result, many of these false alarms were rejected if the signal features changed in a short time (see Fig. 9*b*). When the signal features changed gradually (see Fig. 9*d*), then the algorithm could misinterpret this as a real ABP change. Most false alarms (eight out of nine) were contained in two records (453, 454) relating to one patient. The three false alarms not rejected in the test set contained artifacts that the algorithm could not identify.

The additional ECG–ABP relationship was particularly useful whenever artifacts appeared similar to real physiological changes. Figs 1*d* and 9*c* show episodes of ABP with artifacts similar to the real change in ABP associated with VT/VF, as seen in Fig. 9*f*. Without checking the ECG–ABP relationship, the algorithm would produce an SQI signal that continued to be good in the artifact region, because the algorithm considers this kind of ABP change to be real by looking at the ABP signal alone. Using the ECG–ABP relationship, the algorithm modified the SQI, as shown in Fig. 9*c*, and was able to reject this kind of false alarm without rejecting true alarms, as in Fig. 9*f*. In most

other situations, as in Fig. 7 and the rest of Fig. 9, the ECG–ABP relationship did not significantly affect the SQI.

5 Conclusions

Artifact is a major factor responsible for false alarms. It is useful to derive a signal quality control index to reduce these false alarms. Sometimes, artifacts appear similar to real physiological changes; thus additional information from other related signals is crucial.

This paper presents an approach to reducing false ABP alarms based on both the analysis of ABP signal quality and the use of ECG–ABP relationships. Fuzzy feature representation and reasoning provide a comprehensive and effective way to assess signal quality for each ABP episode. Data suspected of being artifact-corrupted are marked but not discarded. ABP measurements from the data with good SQI appear more reliable than those without adequate signal quality control. By using this approach, false ABP alarms are significantly reduced, with only a small number of rejected true alarms. Note that the number of alarm events missed by the monitors themselves was not determined in this study.

Our results indicate that this approach appears to be effective and practical and should be considered for use in future monitoring systems.

Acknowledgments—This work was supported in part by the Research Resource for Complex Physiologic Signals, NIH/NCRR, grant P41 RR 13622, and by grants from the Hewlett-Packard Research Grants Program and from Nikon Kohden.

The authors would like to thank Thomas Heldt for his review of the manuscript and for his help in preparing the Figures.

Appendix

Table 4 Results of ABP false alarms on development data set

Record number	Length of record, h	Total	TP	TPk	TPr	FP	FPr	FPk
211	21.5	27	27	26	1	0	0	0
212	41.3	29	25	25	0	4	4	0
222	22.4	25	21	21	0	4	4	0
226	31.8	15	5	5	0	10	10	0
230	19.0	19	18	18	0	1	1	0
231	42.0	4	2	2	0	2	2	0
237	42.7	6	0	0	0	6	6	0
252	28.2	6	2	2	0	4	4	0
253	42.5	22	17	17	0	5	5	0
254	42.5	23	11	11	0	12	11	1
262	42.7	3	3	3	0	0	0	0
404	22.3	6	0	0	0	6	6	0
405	22.9	6	0	0	0	6	6	0
410	23.5	18	6	6	0	12	12	0
413	21.4	18	15	15	0	3	3	0
415	42.0	10	10	10	0	0	0	0
451	31.2	26	18	18	0	8	8	0
452	33.7	15	13	13	0	2	2	0
453	46.7	26	17	17	0	9	6	3
454	42.6	42	27	26	1	15	10	5
456	47.2	50	31	31	0	4	4	0
471	62.7	15	8	8	0	7	7	0
472	8.7	7	7	7	0	0	0	0
477	30.0	26	25	25	0	1	1	0
480	19.5	16	11	11	0	5	5	0
Sum	825	445	319	317	2	126	117	9
Average	33	18.4	12.8	12.7	0.1	5.0	4.7	0.4

Total = total alarm annotations from beside monitors; TP = total true positive alarms; TPk = true positives accepted by algorithm; TPr = true positives incorrectly removed by algorithm; FP = total false positive alarms; FPr = false positives removed by algorithm; FPk = false positives remaining after processing of algorithm. PPA gross = 71.69% → 97.24%; PPA average = 64.63% → 98.42%

Table 5 Results of ABP false alarms on test data set

Record number	Length of record, h	Total	TP	TPk	TPr	FP	FPr	FPk
213	49	49	46	46	0	2	2	0
221	24	17	17	17	0	0	0	0
224	47	8	0	0	0	8	8	0
225	46	7	2	2	0	5	5	0
240	30	31	31	31	0	0	0	0
248	34	4	3	3	0	1	1	0
276	58	9	6	6	0	3	2	1

(*continued*)

Table 5 Continued

Record number	Length of record, h	Total	TP	TPk	TPr	FP	FPr	FPk
281	11	4	3	3	0	1	1	0
291	22	4	0	0	0	4	4	0
401	25	8	2	2	0	4	4	0
408	48	21	2	2	0	19	19	0
409	43	40	28	28	0	12	12	0
411	46	16	5	5	0	11	11	0
414	25	45	41	41	0	4	4	0
417	12	9	3	3	0	6	6	0
418	27	3	0	0	0	2	2	0
427	58	53	33	33	0	16	16	0
430	52	1	0	0	0	1	1	0
438	52	12	9	9	0	3	3	0
439	46	27	17	17	0	10	10	0
442	35	5	0	0	0	5	5	0
443	51	20	6	5	1	8	8	0
444	54	38	38	38	0	0	0	0
446	27	48	31	31	0	12	12	0
449	42	37	21	21	0	16	15	1
474	38	10	5	5	0	4	4	0
476	18	33	31	31	0	2	2	0
484	43	66	62	62	0	3	2	1
Sum	1065	604	442	441	1	162	159	3
Average	38	23.0	16.4	16.3	0.0	5.8	5.7	0.1

Total = total alarm annotations from bedside monitors; TP = total true positive alarms; TPk = true positives accepted by algorithm; TPr = true positives incorrectly removed by algorithm; FP = total false positive alarms; FPr = false positives removed by algorithm; FPk = false positives remaining after processing of algorithm. PPA gross = 73.79% → 99.32%; PPA average = 56.72% → 99.24%

References

BERNE, R. M., and LEVY, M. N. (1997): 'Cardiovascular physiology' (Mosby-Year Book, Inc., St. Louis, Missouri, 1997)

CAO, C., KOHANE, I. S., and MCINTOSH, N. (1999): 'Artifact detection in cardiovascular time series monitoring data from preterm infants'. Proc. AMIA Symp., pp. 207–211

CREW, A. D., STOODLEY, K. D. C., LU, R., OLD, S., and WARD, M. (1991): 'Preliminary clinical trials of a computer-based cardiac arrest alarm', *Intens. Care Med.*, **17**, pp. 359–364

FELDMAN, J. M., EBRAHIM, M. H., and BAR KANA, I. (1997): 'Robust sensor fusion improves heart rate estimation: clinical evaluation', *J. Clin. Monit.*, **13**, pp. 379–384

GOLDBERGER, A. L., AMARAL, L. A. N., GLASS, L., HAUSDORFF, J. M., IVANOV, P. C., MARK, R. G., MIETUS, J. E., MOODY, G. B., PENG, C. K., and STANLEY, H. E. (2000): 'PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals', *Circulation*, **101**, pp. e215–e220

LAWLESS, S. T. (1994): 'Crying wolf: false alarms in a pediatric intensive care unit', *Crit. Care Med.*, **22**, pp. 981–985

MAKIVIRTA, A., KOSKI, E., KARI, A., and SUKUVAARA, T. (1991): 'The median filter as processor for a patient monitor limit alarm system in intensive care', *Comput. Methods Programs Biomed.*, **34**, pp. 139–144

MOODY, G. B., and MARK, R. G. (1982): 'Development and evaluation of a 2-lead ECG analysis program', *Comput. Cardiol.*, **9**, pp. 39–44

MOODY, G. B., and MARK, R. G. (1991): 'The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it', *Comput. Cardiol.*, **18**, pp. 185–188

MOODY, G. B., and MARK, R. G. (1996): 'A database to support development and evaluation of intelligent intensive care monitoring', *Comput. Cardiol.*, **23**, pp. 657–660

O'ROURKE, M. F., KELLY, R. P., and AVOLIO, A. P. (1992): 'The arterial pulse' (Lea & Febiger, Malvern, PA, 1992)

PEDRYCZ, W., and GOMIDE, F. (1998): 'An introduction to fuzzy sets: analysis and design' (MIT Press, Cambridge, MA, 1998)

RHEINECK-LEYSSIUS, A. T., and KALKMAN, C. J. (1998): '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', *J. Clin. Monit. Comput.*, **14**, pp. 151–156

TSIEN, C. L., and FACKLER, J. C. (1997): 'Poor prognosis for existing monitors in the intensive care unit', *Crit. Care Med.*, **25**, pp. 614–619

TSIEN, C. L., and KOHANE, I. S. (1998): 'Using multi-signal integration to detect false alarms in intensive care unit', *Med. Decis. Making*, **18**, p. 469

TSIEN, C. L., IOHANE, I. S., and MCINTOSH, N. (2000): 'Multiple signal integration by decision tree induction to detect artifacts in the neonatal intensive care unit', *Artif. Intell. Med.*, **19**, pp. 189–202

WATT, R. C., MASLANA, E. S., and MYLREA, K. C. (1993): 'Alarms and anesthesia', *IEEE Eng. Med. Biol. Mag.*, **12**, pp. 34–41

ZADEH, L. A. (1965): 'Fuzzy sets', *Inform. Control*, **8**, pp. 338–353

ZADEH, L. A. (1983): 'The role of fuzzy logic in the management of uncertainty in expert systems', *Fuzzy Sets Syst.*, **11**, pp. 199–227

ZONG, W., and JIANG, D. (1998): 'Automated ECG rhythm analysis using fuzzy reasoning', *Comput. Cardiol.*, **25**, pp. 69–72

ZONG, W., MOODY, G. B., and MARK, R. G. (1999): 'Reduction of false blood pressure alarms by use of electrocardiogram blood pressure relationships', *Comput. Cardiol.*, **26**, pp. 305–308

ZONG, W., HELDT, T., MOODY, G. B., and MARK, R. G. (2003): 'An open-source algorithm to detect onset of arterial blood pressure pulses', *Comput. Cardiol.*, **30**, pp. 259–262

Author's biography

WEI ZONG received the B.S. degree in electrical engineering and computer science, the MS and the PhD degrees both in Biomedical Engineering, from Xi'an Jiaotong University (XJTU), Xi'an, P. R. China, in 1983, 1986, and 1993 respectively. From 1993 to 1997, he was an associate professor at the Institute of Biomedical Engineering, XJTU. Since 1997, he has been affiliated with the Harvard-MIT Division of Health Sciences and Technology, MIT, the Beth Israel Deaconess Medical Center, and Harvard Medical School, where he is an Instructor in Medicine. His research interests include biomedical signal processing and pattern recognition, physiologic signal databases and medical informatics, and biomedical measurements and instrumentation.