IOP PUBLISHING PHYSIOLOGICAL MEASUREMENT

Physiol. Meas. **29** (2008) 15–32 doi:10.1088/0967-3334/29/1/002

Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter

Q Li[1,2], R G Mark[2,3] and G D Clifford[2,3]

[1] Institute of Biomedical Engineering, School of Medicine, School of Control Science and Engineering, Shandong University, Shangdong, People's Republic of China
[2] Massachusetts Institute of Technology, Cambridge, MA, USA
[3] Harvard–MIT Division of Health Sciences and Technology, Cambridge, MA, USA

E-mail: gari@mit.edu

Received 8 August 2007, accepted for publication 8 November 2007
Published 10 December 2007
Online at stacks.iop.org/PM/29/15

Abstract

Physiological signals such as the electrocardiogram (ECG) and arterial blood pressure (ABP) in the intensive care unit (ICU) are often severely corrupted by noise, artifact and missing data, which lead to large errors in the estimation of the heart rate (HR) and ABP. A robust HR estimation method is described that compensates for these problems. The method is based upon the concept of fusing multiple signal quality indices (SQIs) and HR estimates derived from multiple electrocardiogram (ECG) leads and an invasive ABP waveform recorded from ICU patients. Physiological SQIs were obtained by analyzing the statistical characteristics of each waveform and their relationships to each other. HR estimates from the ECG and ABP are tracked with separate Kalman filters, using a modified update sequence based upon the individual SQIs. Data fusion of each HR estimate was then performed by weighting each estimate by the Kalman filters' SQI-modified innovations. This method was evaluated on over 6000 h of simultaneously acquired ECG and ABP from a 437 patient subset of ICU data by adding real ECG and realistic artificial ABP noise. The method provides an accurate HR estimate even in the presence of high levels of persistent noise and artifact, and during episodes of extreme bradycardia and tachycardia.

Keywords: data fusion, blood pressure, ECG, EKG, electrocardiogram, heart rate, Kalman filter, robust estimation, signal quality

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Physiological signals such as the electrocardiogram (ECG) and arterial blood pressure (ABP) in the intensive care unit (ICU) are often severely corrupted by noise, artifact and missing data, which lead to large errors in the estimation of the heart rate (HR) and ABP (Allen and Murray 1996, Jakob *et al* 2000). This can result in a high incidence of false alarms from ICU monitors, which can sometimes be as high as 90% for some alarm types (Lawless 1994, Tsien and Fackler 1997, Aboukhalil *et al* 2007). Frequent false alarms due to data corruption can also lead to a desensitization of clinical staff to real alarms and a consequent drop in the overall level of care (Chambrin 2001).

Robust HR estimation is essential for ICU monitoring. Beat detection from the ECG is the most direct method for HR measurement (Kohler *et al* 2002), but since ICU noise and artifact are so prevalent, it is difficult for clinicians to believe the monitors' estimates without visual confirmation. Various strategies have been employed to improve estimates of noisy physiological parameters, such as averaging (Jakob *et al* 2000), machine learning (Tsien *et al* 2001), Kalman filtering (Sittig and Factor 1990, Feldman *et al* 1997, Ebrahim *et al* 1997, Tarassenko *et al* 2002, 2003) and signal quality assessment techniques (Allen and Murray 1996, Kaiser and Findeis 2000, Zong *et al* 2004, Chen *et al* 2006). Averaging methods can reduce the influence of transient artifacts, but at the cost of smoothing true physiologic changes. Machine learning techniques detect artifacts efficiently but need large amounts of physiological data to train the model. In contrast to these techniques, Kalman filter (KF) methods have been shown to reliably detect and identify trends, abrupt changes and artifacts from physiological signals with very little knowledge of the underlying model. Furthermore, signal quality assessment methods allow an improved estimate of parameters derived from the recorded data by providing a calibrated reference for identifying periods of high quality data.

HR information can be obtained easily by beat detection from the ECG (Kohler *et al* 2002), pulsatile waveforms such as the ABP (Zong *et al* 2003a) and pulse oximetry waveforms recorded from the photoplethysmogram (Mendelson 1992). These sources provide approximately redundant and independent measures of HR. Furthermore, the sources of the noise and artifact are often weakly correlated or uncorrelated with both the signals that are cardiovascular in origin and with each other. Reliable estimation of HR can therefore be obtained by 'sensor fusion' (Ebrahim *et al* 1997, Feldman *et al* 1997, Tarassenko *et al* 2001), allowing for automatic adjudication of which signal is likely to be a more accurate estimate. Sensor fusion can provide robust cardiovascular parameter estimates even when only one channel of data is relatively noise free. However, if data from an untrustworthy signal are used, the resultant estimate may be degraded. Therefore, methods of determining signal quality and the trustworthiness of the source data are required. In the study presented in this paper, we propose novel signal quality metrics, which we use to adjust a Kalman Filter update, and a modification of Townsend and Tarrasenko's method to fuse these metrics derived from data from multiple sensors (Townsend 2001, Tarassenko *et al* 2002, 2003). This provides a continuously updating estimate of the heart rate that automatically rejects untrustworthy data.

2. Methods

The architecture of our proposed algorithm is based upon a novel integration of a robust KF tracking algorithm and novel signal quality metrics for estimating HR from the ECG and ABP waveforms. After preliminary beat detection, signal quality metrics for the ECG and ABP waveforms are calculated using the combination of several noise metrics. HR is then estimated using a beat detection algorithm and a KF, adjusted to include the signal quality

estimates. Finally, each available estimate of the HR is fused using all available information sources, weighted by the inverse of the Kalman innovation of the signal. Since the innovation is a measure of the novelty in a signal and an artifact is more likely to produce a novel value of the signal source, artifacts tend to be suppressed. However, if more than one information source indicates a change, the innovation rises on multiple sources and allows the algorithm to track the change.

To evaluate our algorithm, we used approximately 6000 h of high quality ICU data with simultaneous ECG and ABP signals, from the Multi-Parameter Intelligent Monitoring for Intensive Care (MIMIC) II database (Saeed *et al* 2002). Since no public database of ECG and ABP with associated signal quality annotations exists, we used stringent thresholds on our novel (but intuitive) signal quality indices (SQIs) to locate long segments of low-noise data. To test the noise sensitivity of our algorithms, real ECG noise taken from the noise-stress test database (Moody *et al* 1984) was added to the ECG. Since no database of real ABP noise exists, we added artificially generated ABP noise to the ABP signal. In order to demonstrate that our algorithm performs well during real episodes of tachycardia and bradycardia, we also tested the method on 2584 human-annotated arrhythmic episodes.

2.1. Signal quality assessment

Signal quality assessment of the ECG was performed by combining four analysis methods: (1) comparison of multiple beat detection algorithms on a single lead, (2) comparison of the same beat detection algorithm on different ECG leads, (3) evaluation of the kurtosis (randomness) of a segment of ECG and (4) calculating the proportion of the spectral distribution of a given ECG segment found to be within a certain physiological frequency band. Signal quality assessment of the ABP was based on a combination of two previously described algorithms: a beat-by-beat fuzzy logic-based assessment of features in the ABP waveform (Zong *et al* 2004) and heuristic thresholding of each ABP pulse to determine normality (Sun *et al* 2006). (These latter algorithms are known as *wsqi* and *jsqi* respectively.)

2.1.1. Comparison of multiple beat detection algorithms on a single lead: bSQI. Since different ECG algorithms are sensitive to different types of noise (Friesen *et al* 1990), the comparison of how accurately multiple QRS detectors isolate each event (such as a beat or noise artifact) provides one estimate of the level of noise in the ECG. In this study, two well-documented open-source QRS detection algorithms with different noise sensitivities were used. One is based on digital filtering (DF) and integration (Hamilton and Tompkins 1986) and other is based on a length transform (LT) after filtering (Zong *et al* 2003b). (These routines are known as *ep_limited* and *wqrs* respectively.) The signal quality of a given ECG lead, with a window of length w seconds, is defined to be the ratio of beats detected synchronously (within an interval, γ) by both algorithms to all the detected beats (by either algorithm) within the window (Oefinger 2006). In this study, w is set to 10 s to be commensurate with the window over which the HR is calculated (see section 3.1) and γ is set to be 150 ms as per the recommendations of the American National Standards Institute (ANSI/AAMIEC57 1998). The DF and LT techniques have been shown to achieve a sensitivity of 99.69% and 99.65% respectively, and both had a positive predictivity of 99.77% (Hamilton and Tompkins 1986, Zong *et al* 2003b) when evaluated on the MIT/BIH arrhythmia database (Goldberger *et al* 2000, Moody and Mark 2001). However, QRS positive predictivity of the LT technique drops markedly when the signal-to-noise ratio (SNR) is low, since the LT is sensitive to noise. A consensus beat detection signal quality index (*bSQI*) was defined for the kth beat as

$$bSQI(k) = N_{\text{matched}}(k, w)/N_{\text{all}}(k, w) \tag{1}$$

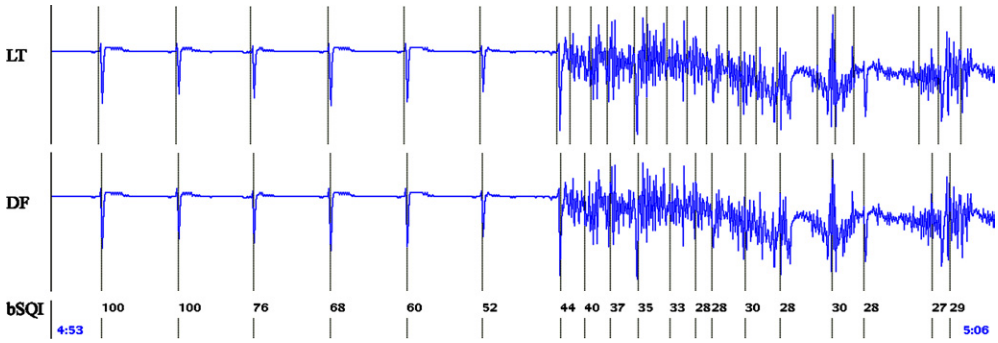


Figure 1. The *bSQI* of a 13 s ECG segment that is clean for the first half of the analysis segment and then is noisy for the second half of the segment. The same lead is shown twice, with vertical fiducial markers for each beat detected by either the LT technique (upper plot) or DF technique (lower plot). Note that the SQI for each beat is derived from a 10 s window, centered on each beat, and so the SQI begins to drop 5 s before the ECG becomes noisy.

where N_{matched} is the number of beats that both algorithms agreed upon (within $\gamma = 150$ ms) and N_{all} is the number of all beats detected by either algorithm (without double counting the matched beats). In other words, $N_{\text{all}} = N_{\text{DF}} + N_{\text{LT}} - N_{\text{matched}}$, where N_{DF} is the number of beats detected by the Hamilton and Tompkins (DF) method and N_{LT} is the number of beats detected by the LT method. *bSQI* therefore ranges between 0 and 1 inclusively. For N beats, there are N windows set to be $w = 10$ s long, centered $\pm$ 5 s around the kth beat. Figure 1 illustrates the calculation of the *bSQI* (multiplied by 100). Note that each beat was correctly detected by both algorithms (indicated by vertical lines), and no erroneous detections were made until half way through the segment. The *bSQI* then drops significantly (with an overlap into the clean segment due to the 5 s trailing window). Note also that a *bSQI* value is only given when the DF technique reports a beat, since this is the least noise sensitive of the two techniques (Friesen *et al* 1990).

2.1.2. Beat detection comparison using different ECG leads: iSQI. In theory, when synchronous ECG leads are available, comparison between different leads can provide more accurate estimates of HR and signal quality, since more information is present. Although we can calculate *bSQI* for different leads separately, if one QRS detection algorithm misses one or more beats (due to low QRS amplitudes) or registers extra beats (due to artifact or high amplitude T waves), *bSQI* will fail to give a good signal quality measure. Figure 2(a) shows erroneous extra beat detection by the LT method and correct beat detection by the DF method. In such a situation, the *bSQI* is low, yet the DT method is obviously providing a good estimate of the heart rate. Beat detection comparison between different leads, using the same QRS detection algorithm, allows us to see that the SQI is in fact high. An inter-channel signal quality index (*iSQI*) was calculated as the ratio of the number of matched beats (N_{matched}) to all detected beats (N_{all}) between a given lead and all other synchronous ECG leads, using only the DF method. Subsequently, the maximum value over each 10 s epoch ($\pm$ 5 s around the current beat) is calculated for each beat. That is,

$$iSQI(k)_i = \max(N_{\text{matched}}(k, w)_{i,j} / N_{\text{all}}(k, w)_{i,j}) \qquad \forall j,\ j \neq i \tag{2}$$

where i is the current ECG lead and j represents each of the other different leads. Figure 2(b) illustrates how *iSQI* works even when *bSQI* fails. *iSQI* is effective when more than one ECG lead has good signal quality.

Figure 2. The *bSQI* and *iSQI* of an ECG segment. (a) *bSQI* of lead V fails because extra beats are erroneously detected by the LT method; (b) *iSQI* works by comparing detection results on lead V with those from lead MCL1.

2.1.3. Kurtosis of the ECG: kSQI. From the central limit theorem we know that random uncorrelated processes, such as thermal noise, tend to have Gaussian distributions. Conversely, correlated signals tend to have non-Gaussian distributions. A simple measure of how Gaussian-like a signal appears to be is kurtosis, the fourth standardized moment of a distribution, which measures the relative peakedness of a distribution with respect to a Gaussian distribution. The kurtosis, K, of signal x with mean μ_x and standard deviation σ is defined as $K = E\{(x-\mu_x)^4\}/\sigma^4$ where $E\{\}$ is the mathematical expectation operator. The empirical estimate of kurtosis, $\hat{K}$, of a discrete signal x_i is given by

$$\hat{K} = \frac{1}{M}\sum_{i=1}^{M}\left[\frac{x_i - \hat{\mu}_x}{\hat{\sigma}}\right]^4 \tag{3}$$

where $\hat{\mu}_x$ and $\hat{\sigma}$ are the empirical estimate of the mean and standard deviation of x_i respectively, and M is the number of samples in the dataset. The kurtosis of a Gaussian distribution is equal to 3 and clean, sinus rhythm ECG generally has a kurtosis larger than 5 (He *et al* 2006). Muscle artifact has a kurtosis around 5 and baseline wander and power-line interference have kurtosis lower than 5 (Clifford 2006). The kurtosis of an ECG segment was calculated for 10 s epochs (centered $\pm$ 5 s around the current beat) and the kurtosis-based SQI (*kSQI*) is defined by

$$kSQI(k) = \begin{cases} 1 & \text{if} \quad \text{kurtosis } (k, w) > 5 \\ 0 & \text{if} \quad \text{kurtosis } (k, w) \leqslant 5 \end{cases}. \tag{4}$$

Low *kSQI* usually indicates low frequency noise such as baseline wonder, Gaussian (thermal observation) noise and high frequency sinusoidal noise (power-line interference).

2.1.4. Spectral distribution of ECG: sSQI. Since the QRS energy is mainly concentrated in a 10 Hz wide frequency band, centered around 10 Hz (Murthy *et al* 1978), the ratio of the power spectral density (PSD) in this band compared to the PSD in the overall signal provides

a measure of the signal quality. The spectral distribution ratio (*SDR*) of an ECG segment was defined to be the ratio of the sum of the power, P, of the ECG between frequencies, f, of 5 Hz and 14 Hz to the power between 5 Hz and 50 Hz as follows:

$$SDR(k) = \int_{f=5}^{f=14} P(k, w)\,\mathrm{d}f \Big/ \int_{f=5}^{f=50} P(k, w)\,\mathrm{d}f. \tag{5}$$

When the *SDR* is low, high frequency noise contamination in the ECG, such as muscle artifact, is likely. When the *SDR* is extremely high, an increased presence of QRS-like artifact, such as electrode motion, is likely. Moderate values of *SDR* indicate good ECG quality. Thus, the spectral distribution signal quality index (*sSQI*) is defined as

$$sSQI(k) = \begin{cases} 1 & \text{if} \quad SDR \geqslant 0.5 \quad \text{and} \quad SDR \leqslant 0.8 \\ 0 & \text{if} \quad SDR < 0.5 \quad \text{or} \quad SDR > 0.8. \end{cases} \tag{6}$$

These thresholds were empirically determined through repeated observations and can be adjusted slightly with little effect.

2.1.5. Combined ECG SQI. The combined ECG signal quality index (*ECGSQI*) is derived by combining the four SQI metrics detailed above as follows:

$$ECGSQI(k) = \begin{cases} \max(bSQI(k), iSQI(k)) & \text{if} \quad kSQI(k) = 1 \ \& \ sSQI(k) = 1 \\ bSQI(k) & \text{if} \quad kSQI(k) = 1 \ \& \ sSQI(k) = 0 \\ \max(bSQI(k), iSQI(k))^{*}\eta & \text{if} \quad kSQI(k) = 0 \ \& \ sSQI(k) = 1 \\ bSQI(k)^{*}\eta & \text{if} \quad kSQI(k) = 0 \ \& \ sSQI(k) = 0 \end{cases} \tag{7}$$

where η is a positive coefficient less than unity that indicates the presence of spectral or statistical noise. In this implementation we chose $\eta = 0.7$, reducing the SQI by 30% if such noise is detected. However, the actual value of this coefficient is fairly arbitrary, since a value for the overall SQI must be tuned to a particular application. In effect, η can be thought as the boundary between moderate and high quality data.

The rationale for the logic in (7) is as follows: if *sSQI* and *kSQI* indicate good quality ECG, we can trust the *bSQI* and *iSQI* metrics, and therefore we set the *ECGSQI* to be the maximum of these two metrics. If, however, *sSQI* is low, out-of-band noise is present and we can only trust *bSQI*. (We do not trust *iSQI* in this situation because artifact can often be present on both channels and result in matched detections of artifact on both channels. However, this artifact is not likely to be spectrally coherent and will present a low value of *sSQI*.) If *kSQI* indicates the signal is of low quality, we follow the same logic, but we trust *bSQI* or *iSQI* less (by a multiplicative factor, η). Figure 3 illustrates the process of deriving an *ECGSQI* value for a particular 13 s segment of two-lead ECG (MCL1 and AVF leads). Note that the metric is calculated for a $\pm$ 5 s window around each beat. The second and third traces are the outputs from the two QRS detectors (LT and DF) applied to the MCL1 lead (denoted by a subscripted '1'). The upper trace (DF_2) is the output of the digital filter QRS detector on the other lead (AVF). The lower five traces represent the different SQIs described above. The fourth trace comprises vertical markers for each of the consensus beat detections and the associated *bSQI* on the MCL1 lead. The consensus of DF_1 and DF_2 (the inter-lead signal quality, *iSQI*) is given on the fifth trace. The sixth trace illustrates the output of the kurtosis-based metric (*kSQI*) and the seventh trace the effect of the spectral thresholding metric, *sSQI*, on lead MCL1. The resultant combined *ECGSQI* metric is shown on the final (lower) trace. Note that although each metric indicates the noisy portion of the signal has lower SQI, each metric contributes to a lowered overall *ECGSQI* during the noisy period. Note also that since the *bSQI* and *kSQI*

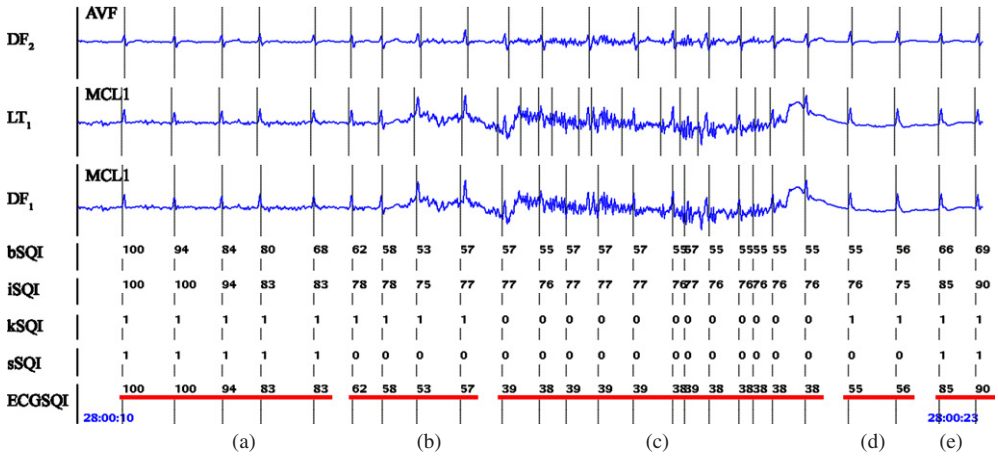


Figure 3. An example of the calculation of *ECGSQI*. *bSQI*, *kSQI* and *sSQI* are based on lead MCL1 and *iSQI* is based on leads MCL1 and AVF. Segments (a) and (e) possess good *kSQI*, *sSQI* and *ECGSQI*; segments (b) and (d) possess good *kSQI* but bad *sSQI* and medium *ECGSQI*; segment (c) possesses low *kSQI* and *sSQI*, and hence low *ECGSQI*. Note the ECGSQI illustrated here is relative to the MCL1 lead.

are calculated for the MCL1 lead in figure 3, the *ECGSQI* illustrated is relative to that lead, as it is the noisiest lead. If the *bSQI* and *kSQI* for the AVF lead were used, the *ECGSQI* would be a little higher, since the AVF lead is less noisy. It is actually the higher *ECGSQI* that is usually reported.

2.1.6. Signal quality assessment of ABP. Two previously developed ABP signal quality assessment methods, *wSQI* (Zong *et al* 2004) and *jSQI* (Sun *et al* 2006), were employed to obtain the SQI of ABP. The *wSQI* algorithm consists of an ABP pulse detection routine, using an open-source ABP onset detection algorithm, *wabp* (Zong *et al* 2003a), a waveform feature extraction routine, a waveform feature fuzzy representation and a fuzzy reasoning procedure to produce the SQI. This algorithm was previously trained on data from the MIMIC DB (Moody and Mark 1997) and has been shown to give an accurate assessment of ABP signal quality in previous studies (Zong *et al* 2004). A *wSQI* value is associated with each beat and possesses a continuous value between 0 and 1 (poor to excellent quality). Values of *wSQI* above 0.5 corresponded to good signal quality, where reliable heart rate and blood pressure estimates can be made. The *jSQI* algorithm uses the same beat detection algorithm (a direct translation from the C programming language to Matlab) and identifies a series of features in each ABP pulse. Plausible heuristic constraints are set on the ABP amplitudes, slopes and beat-to-beat variations in each pulse in order to generate a signal abnormality index, *jSQI*, which takes a binary value: 0 for normal beats and 1 for abnormal beats (physiologically abnormal or noise/artifact).

The ABP signal quality index (*ABPSQI*) is calculated by combining *wSQI* and *jSQI* as follows:

$$ABPSQI(k) = \begin{cases} wSQI & \text{if} \quad jSQI = 0 \\ wSQI^{*}\eta & \text{if} \quad jSQI = 1 \end{cases} \tag{8}$$

where $1 \geqslant \eta \geqslant 0$ is the positive coefficient chosen to be $\eta = 0.7$, i.e. if *jSQI* indicated a good quality signal, *wSQI* can be believed. Otherwise, *wSQI* is trusted less and is therefore

multiplied by the coefficient η. Although this coefficient may differ from the analogous coefficient in (7), we chose equivalent values, for consistency in interpretation.

2.2. Kalman filtering for HR and ABP tracking

2.2.1. Kalman filtering algorithm. The KF is an optimal state estimation method for a stochastic signal (Brown 1983, Welch and Bishop 2004) that estimates the state of a discrete-time controlled process, x, with measurement data z, where x and z are governed by the linear stochastic difference equations

$$x_k = \boldsymbol{A}x_{k-1} + \boldsymbol{B}u_k + w_{k-1}, \tag{9}$$

$$z_k = Hx_k + v_k. \tag{10}$$

The random variables w and v are independent, white, and possess normal probability distributions, $p(w) \sim N(0, Q)$ and $p(v) \sim N(0, R)$. The matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{H}$ are the coefficient state transition matrices, Q is the state noise covariance, $\boldsymbol{R}$ is the measurement noise covariance and u is an optional control input to the state x.

The KF algorithm is given by the following equations:

$$\hat{x}_k^- = \boldsymbol{A}\hat{x}_{k-1} + \boldsymbol{B}u_k \tag{11}$$

$$\boldsymbol{P}_k^- = \boldsymbol{A}P_{k-1}\boldsymbol{A}^T + \boldsymbol{Q} \tag{12}$$

$$\boldsymbol{K}_k = \boldsymbol{P}_k^-\boldsymbol{H}^T(\boldsymbol{H}P_k^-\boldsymbol{H}^T + \boldsymbol{R})^{-1} \tag{13}$$

$$\hat{x}_k = \hat{x}_k^- + \boldsymbol{K}_k(z_k - \boldsymbol{H}\hat{x}_k^-) \tag{14}$$

$$\boldsymbol{P}_k = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H})\boldsymbol{P}_k^- \tag{15}$$

where $\hat{x}_k^-$ and $\hat{x}_k$ are *a priori* and *a posteriori* state estimate before and after a given measurement z_k, $\boldsymbol{P}_k^-$ and $\boldsymbol{P}_k$ are the error covariance of *a priori* and *a posteriori* estimate, $r_k = z_k - \boldsymbol{H}\hat{x}_k^-$ is the measurement innovation (or residual) and $\boldsymbol{K}_k$ is the gain required to minimize the *a posteriori* error covariance, $\boldsymbol{P}_k$.

We implemented a KF to estimate the HR derived from ECG and ABP separately and the systolic, mean and diastolic blood pressure derived from ABP (Zong *et al* 2004). The performance of the algorithm on blood pressure estimation will be reported in a separate article (Li *et al* 2007). Combining the outputs from each KF was performed in a similar manner to Townsend and Tarassenko (Townsend 2001, Tarassenko *et al* 2002, 2003). However, in order to more heavily weight estimates derived from cleaner data, we propose the use of the *SQI* (either *ECGSQI* or *ABPSQI* depending on the source) to adjust the measurement noise covariance, $\boldsymbol{R}$, when $\boldsymbol{K}_k$ is updated. When the *SQI* is low, z_k should be trusted less, so $\boldsymbol{K}_k$ should be small, and hence we force $\boldsymbol{R}$ to be large. This is achieved by modifying $\boldsymbol{R}$ as follows:

$$\boldsymbol{R} \rightarrow \boldsymbol{R} \cdot \exp(1/SQI^2 - 1). \tag{16}$$

This nonlinear weighting function therefore tends to unity as the *SQI* tends to unity and so does not affect the measurement noise covariance. This has the effect of forcing the KF to trust the current measurement, z_k, and elevating the Kalman gain, $\boldsymbol{K}_k$. At low SQIs $\boldsymbol{R}$ tends to infinity (but in practice is limited to a large value) and forces the KF to reduce $\boldsymbol{K}_k$ and hence trust the previous measurements more. Furthermore, an upper limit that defines the cusp between good and bad data, SQI_{th}, is defined. When $SQI < SQI_{\text{th}}$, the KF is not updated. The determination of the value of SQI_{th} is described in section 3.1.

2.2.2. KF initialization and operation. Following Tarassenko *et al* (2002), we pick the simplest form of the KF, and set the state to be a scalar, the heart rate. (Each matrix associated with the KF is now denoted in plain face to reflect that they are scalars.) The heart rate can be calculated in many ways and is expressed as an average number of beats per minute (bpm). We considered three methods: (a) the number of detected beats in a given window (scaled to one minute), (b) 60 divided by the mean RR (peak-to-peak) interval in a given window and (c) 60 divided by the median RR interval in a given window. We used a window length of 10 s, which is typical in clinical practice. Furthermore, the heart rate at each moment is assumed to be approximately equal to the heart rate at the next moment ($A \approx 1$). After neglecting the control input, (11) then reduces to $\hat{x}_k^- = \hat{x}_{k-1}$. In order to initialize the KF, one must estimate Q, the state noise covariance matrix, and R, the measurement noise covariance, to calculate P_k^- and K_k. R was similarly initialized to unity, noting that it is immediately modified by the SQI to reflect our trust in the data. Q was empirically adjusted to have an initial value of $Q = 0.1$. Values of $Q \ll 0.1$ lead to the KF trusting the state estimate too much and not adapting to the new initial observations. Values of $Q \gg 0.1$ lead to the KF trusting the new observations too much and simply following new HR observations too closely. Setting H to unity then allows us to estimate the Kalman gain, K_k, from (13) and hence the *a posteriori* error covariance estimate, P_k, from (15). The filter can then be run online with only a few iterations (heart beats) for convergence. The Kalman residual is then given as $r_k = z_k - \hat{x}_k^-$ at each update (each detected beat).

2.3. Data fusion for HR estimation

Once a Kalman filtered HR signal has been derived from the ECG and ABP signals separately, they must be combined to derive a robust consensus estimation of HR. By modifying the approach of Townsend and Tarassenko (Townsend 2001, Tarassenko *et al* 2002, 2003), we use the SQI-weighted residual error, r, of each KF:

$$HR = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} HR_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} HR_2 \tag{17}$$

where HR_1 is the heart rate derived from the ECG and HR_2 is derived from the ABP, and the SQI-scaled innovations are given by $\sigma_1^2 = (r_1/SQI_1)^2$ and $\sigma_2^2 = (r_2/SQI_2)^2$. In this way, when one channel (e.g. channel 1) is corrupted by artifact and the HR (HR_1) is miscalculated, the SQI (SQI_1) will be low and the sudden change of HR (HR_1) will make the residual error (r_1) large. The weighted innovation $\left(\sigma_1^2\right)$ will therefore be large and the weighting for HR_1, $\sigma_2^2 \big/ \left(\sigma_1^2 + \sigma_2^2\right)$, will be small. The estimation of the HR will then rely more on HR_2 than HR_1.

There are four general scenarios for our robust estimation-based algorithm:

(1) *Normal, stable HR.* Low residual error and high *SQI* on both channels, so that both HR_1 and HR_2 are weighted approximately equally and the fused HR can be trusted.
(2) *Large, but physiologically reasonable changes of HR.* A high residual error and high *SQI* on both channels. Even though the residual may be relatively high, it is high for both channels, both HR_1 and HR_2 are again weighted approximately equally and the fused HR can be trusted.
(3) *Artifact on one channel.* A low *SQI* and high residual error (and sometimes low depending on the artifact type) on one channel only. The HR information of this channel will be ignored because of the low SQI weighting and the fused HR can be trusted.
(4) *Artifact on both channels.* HR information is corrupted on both channels leading to a low *SQI* on both channels. In this case, the current HR should not be trusted and the KF makes a guess based upon the previous HR values and current (learned) Kalman gain.

It should be noted that this fusion estimation can be extended to an arbitrary number of n channels for any parameter X as follows (Townsend 2001):

$$X = \sum_{k=1}^{n} \left(\frac{\prod_{i=1, i\neq k}^{n} \sigma_i^2}{\sum_{i=1}^{n} \left(\prod_{j=1, j\neq i}^{n} \sigma_j^2 \right)} \cdot X_k \right) \qquad k = 1, 2, \ldots, n. \tag{18}$$

This formulation is particularly useful for the ICU data where multiple estimates of the same physiological parameter can be derived. For example, one might use the photoplethysmogram, or pulmonary arterial pressure, as well as the ABP and ECG to determine physiological parameters such as HR, ABP or cardiac output.

2.4. Evaluation data

2.4.1. Clean normal data. The HR fusion algorithm was evaluated on a large subset of the MIMIC II database (Saeed *et al* 2002). The following criteria were used to determine low-noise segments of the database: *ECGSQI* $\geqslant$ *0.95 and ABPSQI* $\geqslant$ *0.95 and* $\geqslant$ *1 h duration*, with the ABP and at least one channel of ECG simultaneously present. From the 2500 patients comprising a total of 150 000 h of simultaneous ECG and ABP data, the clean dataset included 437 subjects, comprising 3762 1 h or longer (1.62 ± 0.69 h) data segments or 6084 h in total. (This illustrates how rare it is for an ICU patient to have 1 h of undisturbed physiological data.) ECG noise and ABP noise were separately added to the clean dataset at different SNRs to generate the noisy evaluation dataset. It should be noted that previous studies have shown that artifact occurs more frequently when there is a change in the heart rate (Clifford *et al* 2002), and therefore our data are likely to be biased toward non-changing heart rates. However, if we analyze the distribution of true HR range (maximum HR–minimum HR) for each segment of the clean dataset we find that a large number of test segments differ by more than 10 or even 20 bpm (a significant change in HR). The distribution is approximately log-normal with a mean of 13.3 bpm and a standard deviation of 8.9 bpm.

2.4.2. Additive noise. The ECG noise introduced included electrode motion artifact (*EM*), baseline wander (*BW*) and muscle artifact (*MA*), each with two channels of simultaneously recorded data, taken from PhysioNet's NSTDB database (Moody *et al* 1984, Goldberger *et al* 2000). [Since no ABP noise-stress test database exists, realistic artificial ABP artifacts were used to test the ABP. The description of these artifacts is extensive and the analysis will be presented in a forthcoming paper (Li *et al* 2007).] If more than two channels of ECG are present, the first channel of the NSTDB is added to the odd channels and the second channel of the NSTDB is added to the even channels.

Each of the three ECG noise types (*MA*, *BW*, *EM*) were added separately to each ECG lead of the clean dataset of varying SNRs, for every other 5 min epoch in the clean data (giving six noisy 5 min periods, each followed by 5 min of clean data in each hour). The SNR during the noisy segments was set to a value of 24, 18, 12, 6, 0 and -6 dB separately, giving a total of six different datasets (with different SNRs). Figure 4 illustrates the clean data and noisy ECG data with differing SNR levels.

2.4.3. Abnormal data. As mentioned above, HR estimation is extremely important as a first-order estimate of cardiovascular system performance. However, the exact value of the HR depends on the size of the window over which it is calculated. Although a multi-scale HR metric may be more appropriate, it is not as easy to interpret, and we have therefore used the clinical convention of measuring the HR over (non-overlapping) 10 s

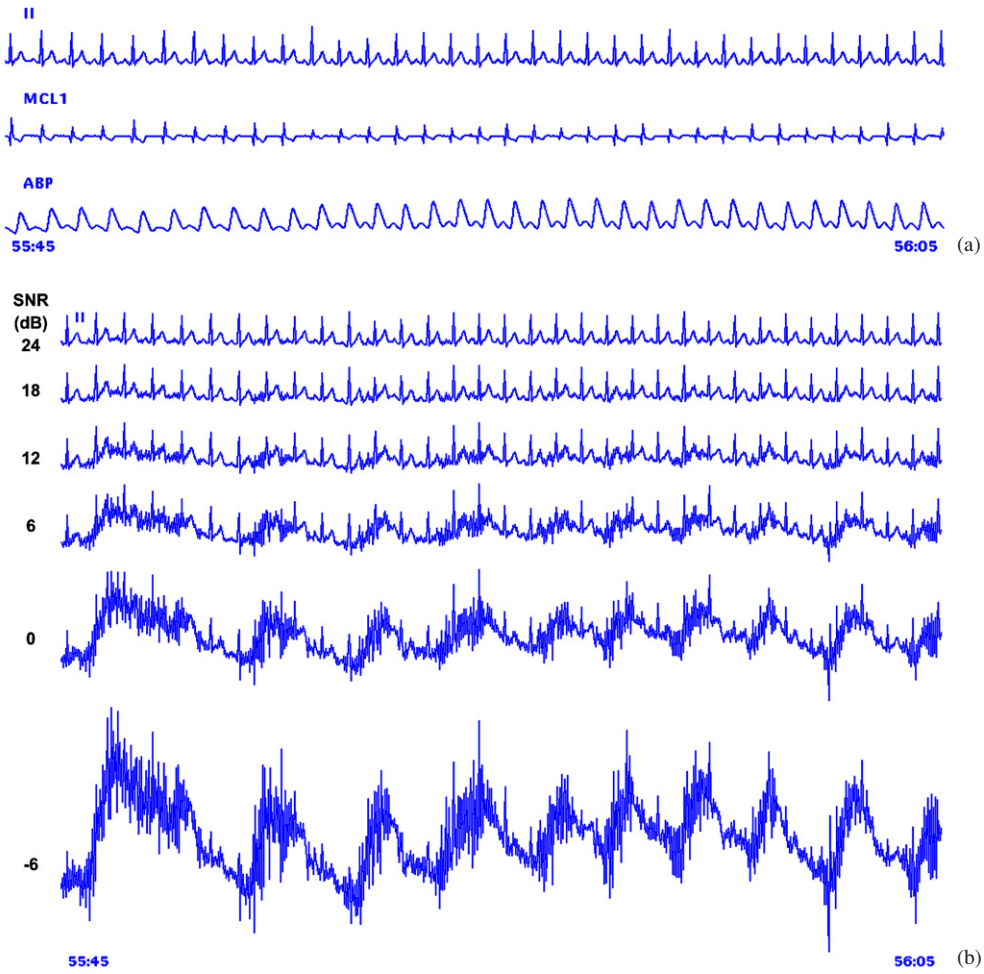
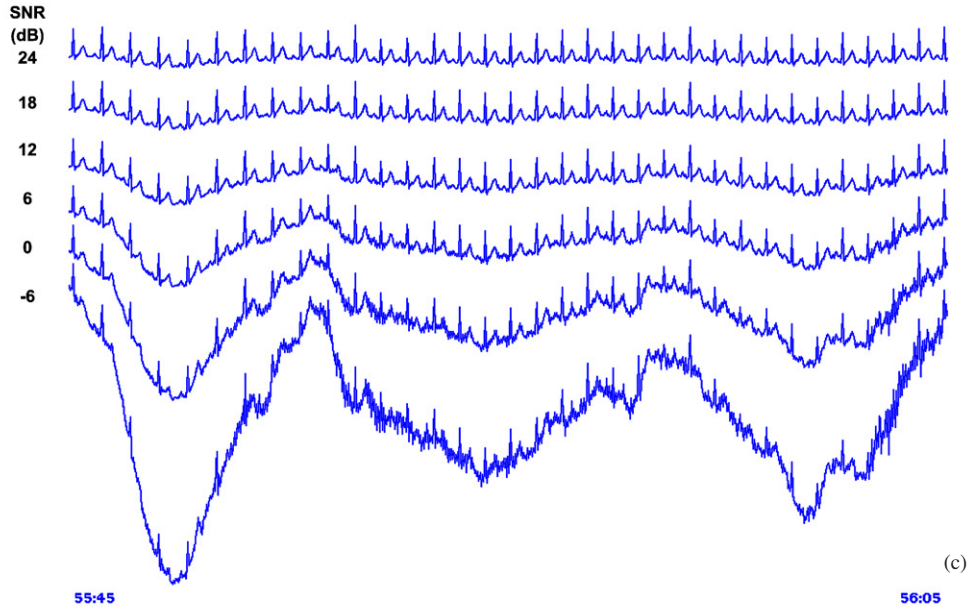


Figure 4. An example of 20 s of evaluation data. (a) Clean data with two leads of ECG (lead II and MCL1) and one ABP signal; (b) *MA* noise at different SNRs; (c) *BW* noise at different SNRs; (d) *EM* noise at different SNRs (SNR = 24, 18, 12, 6, 0 and −6 dB).

epochs. However, arrhythmias may manifest over time scales much shorter than this (often 5 s or less). It is therefore important to ensure that the algorithm works on both normal and abnormal rhythms and hence we require another dataset on which to test our algorithm. In a related piece of work (Clifford *et al* 2006, Aboukhalil *et al* 2007), we developed a false alarm suppression algorithm for the ICU using a subset of over 5500 life-threatening alarms (asystole, bradycardia, tachycardia, ventricular tachycardia and ventricular flutter/fibrillation) taken from the same MIMIC II database, for which we have 45,000 h of simultaneous ECG and ABP data. This includes 707 episodes of bradycardia (of which 506 are true and 201 false) and 1877 episodes of tachycardia (of which 1444 true and 433 false). These false alarm rates are typical of alarms in the ICU, which can be as high as 90% (Aboukhalil *et al* 2007), but for life-threatening alarms are usually around 40% (Aboukhalil *et al* 2007). Therefore, this subset of the MIMIC II data provides an excellent test set for the HR estimation algorithms. We do not consider the algorithm presented in this paper to be applicable to asystole, since



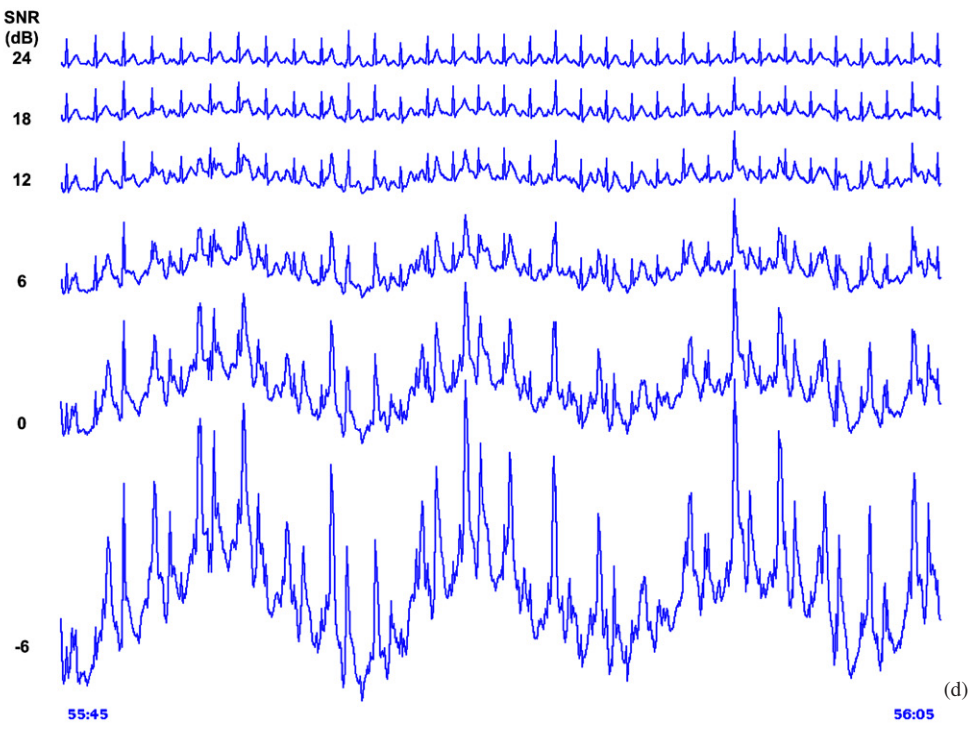


Figure 4. (Continued.)

asystoles shorter than 10 s would not be detected. Furthermore, ventricular arrhythmias are waveform morphology related, and therefore are not relevant to the HR-tracking algorithm presented here. If we consider 20 s epochs around each bradycardia and tachycardia alarm

(with the alarm occurring at 17 s in the epoch) then we are able to construct a new test set of 2584 events (1950 true episodes) and over 14 h (861 min) of simultaneous ECG and ABP with episodes of significant changes in the heart rate.

2.5. Gold standard HR calculation method

Following standard clinical practices, we calculate the HR (and SQI) for each (non-overlapping) 10 s epoch of ECG and ABP. We have shown that the Kalman filter-based fusion approach (equation 17) described in this work leads to an accurate estimate of HR with an average absolute error of 0.02 ± 0.12 bpm and zero bias (Clifford and Li 2007). This modified Kalman fused estimate is therefore used as the 'Gold Standard' HR by which to evaluate the method under noisy conditions.

3. Results

3.1. HR estimation for varying ECG noise levels

The HR is easily calculated from a clean electrocardiogram (ECG) or blood pressure (BP) signal by counting the number of pulses or wave packets representing each beat that are observed in 1 min. However, HR is generally measured by counting beat intervals over smaller windows than 60 s (typically 10–30 s) and the result is scaled to units of bpm by multiplying by $60/T$, where T is the size of the temporal window in units of seconds. To simulate current clinical practices and evaluate the estimation method presented in this paper, seven HR estimation methods were chosen for evaluation, namely:

(1) HR estimation from single channel QRS detector (DF or LT),
(2) HR estimation from single channel QRS detector using sample-and-hold when $SQI < SQI_{\text{th}}$ (SH1) [This simulates the operation of single channel ICU equipment behavior],
(3) Sample-and-hold combined with HR estimate selected at each epoch from the leads with the highest SQI (SHm) [to simulate multichannel ICU equipment behavior],
(4) HR estimation using the Kalman filter and SQI (KF),
(5) HR estimation from ABP beat analysis (*wabp*),
(6) HR estimation using the Kalman filter and SQI from ABP beat analysis (KFABP),
(7) HR estimation by fusing ECG and ABP HR estimates after KF (FUSE).

To determine a reasonable value of SQI_{th} for the sample-and-hold and Kalman filter algorithms, we removed the Kalman gain (K_k) update and SQI threshold control at the Kalman filter step. Figure 5 illustrates the root mean squared error (rMSE) of the difference between the Kalman filtered HR and 'Gold standard' HR. The error becomes larger when the signal quality is low. Although a value of $SQI_{\text{th}} = 0.7$ appears to be a good choice of SQI_{th}, a lower value is actually appropriate. This is because, if the SQI_{th} is set too high, the Kalman filter will stop updating on some physiologically real HR changes. Through trial and error we selected $SQI_{\text{th}} = 0.5$.

Table 1 provides the mean SQI for each HR estimation method averaged over the whole evaluation dataset. Average SQI is given for each of the three different types of noises and over the entire range of SNRs considered.

3.2. HR estimation during bradycardia and tachycardia

To determine whether the HR estimation algorithm correctly tracks changes in the heart rate under extreme circumstances, we evaluated whether our algorithm accurately tracked HR during

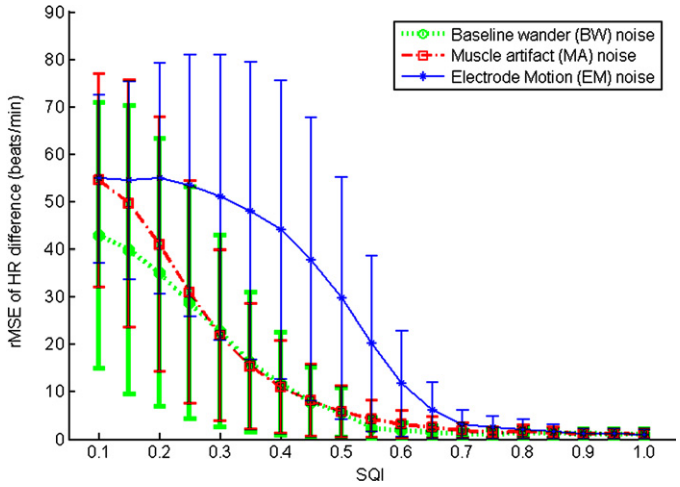


Figure 5. Root mean squared error (rMSE) of the HR difference between the Kalman filtered HR and 'Gold standard' HR without SQI control.

Table 1. *ECGSQI* and HR estimation error for different types of noise and SNR.

Noise type	SNR (dB)	ECGSQI	HR (LT) rMSE (bpm)	HR (DF) rMSE (bpm)	HR (SH1) rMSE (bpm)	HR (SHm) rMSE (bpm)	HR (KF) rMSE (bpm)	HR (FUSE) rMSE (bpm)
MA	24	1.00 ± 0.04	3.1 ± 2.8	2.8 ± 2.4	1.8 ± 1.1	1.9 ± 1.1	1.1 ± 0.9	0.9 ± 0.8
	18	0.97 ± 0.08	3.8 ± 3.5	2.8 ± 2.4	1.9 ± 1.3	1.9 ± 1.1	1.1 ± 0.9	0.9 ± 0.8
	12	0.66 ± 0.30	35.3 ± 30.3	3.2 ± 2.8	3.0 ± 2.7	2.2 ± 1.6	1.6 ± 1.4	1.0 ± 0.8
	6	0.22 ± 0.28	80.7 ± 48.7	19.0 ± 17.0	3.2 ± 2.8	2.5 ± 2.0	2.1 ± 1.8	1.0 ± 0.9
	0	0.08 ± 0.19	105.6 ± 46.9	42.7 ± 29.7	3.9 ± 3.6	2.6 ± 2.0	2.3 ± 2.0	1.0 ± 0.8
	−6	0.03 ± 0.10	123.1 ± 38.9	51.1 ± 30.1	4.0 ± 3.7	2.6 ± 2.0	2.4 ± 2.0	1.0 ± 0.8
BW	24	0.98 ± 0.07	3.09 ± 2.71	2.8 ± 2.4	1.8 ± 1.2	1.91 ± 1.13	1.1 ± 0.9	0.9 ± 0.8
	18	0.90 ± 0.12	3.11 ± 2.72	2.8 ± 2.4	1.8 ± 1.2	1.91 ± 1.13	1.1 ± 0.9	0.9 ± 0.8
	12	0.78 ± 0.12	3.13 ± 2.75	2.8 ± 2.4	1.9 ± 1.2	1.91 ± 1.13	1.1 ± 1.0	1.0 ± 0.8
	6	0.70 ± 0.08	4.32 ± 4.00	2.8 ± 2.4	1.9 ± 1.3	1.93 ± 1.16	1.3 ± 1.0	1.0 ± 0.8
	0	0.64 ± 0.15	21.24 ± 20.44	4.8 ± 4.5	1.9 ± 1.4	2.00 ± 1.26	1.4 ± 1.2	1.0 ± 0.8
	−6	0.49 ± 0.27	50.19 ± 44.48	16.8 ± 15.8	2.8 ± 2.4	2.82 ± 2.33	2.5 ± 2.3	1.0 ± 0.8
EM	24	1.00 ± 0.04	3.04 ± 2.66	2.8 ± 2.4	1.8 ± 1.2	1.92 ± 1.13	1.1 ± 0.9	0.9 ± 0.8
	18	0.98 ± 0.07	3.35 ± 2.97	2.8 ± 2.4	2.0 ± 1.5	1.92 ± 1.13	1.1 ± 0.9	0.9 ± 0.8
	12	0.79 ± 0.25	18.21 ± 16.28	9.2 ± 8.8	3.0 ± 2.6	2.36 ± 1.77	1.6 ± 1.4	1.0 ± 0.8
	6	0.44 ± 0.30	53.81 ± 36.29	39.0 ± 30.1	5.4 ± 5.1	5.35 ± 4.96	3.8 ± 3.5	1.1 ± 0.9
	0	0.13 ± 0.23	80.13 ± 33.84	57.5 ± 31.3	3.5 ± 3.1	4.53 ± 4.10	2.9 ± 2.6	1.0 ± 0.8
	−6	0.04 ± 0.13	94.31 ± 36.71	54.6 ± 27.4	4.0 ± 3.7	3.80 ± 3.38	2.6 ± 2.3	1.0 ± 0.8

episodes of paroxysmal bradycardia and tachycardia. Based on previous work (Clifford *et al* 2006), we calculated heart rates using the median of the four shortest RR intervals for tachycardic episodes, and the four longest RR intervals for bradycardic episodes to ensure rapid changes could be estimated. Of the 707 bradycardia alarms (506 true and 201 false), our algorithm correctly tracked HR in 167 of the 201 false alarms and 500 of the 506 true alarms. In other words, the HR estimation algorithm correctly tracked the true abnormal drops in HR 99% of the time and was only 'fooled' into tracking the artifacts (that tricked the monitors into alarming) 17% of the time. In the case of tachycardia, HR was estimated inaccurately in

Table 2. HR estimation error with and without SQI control.

SNR (dB)	MA with SQI rMSE (bpm)	MA without SQI rMSE (bpm)	BW with SQI rMSE (bpm)	BW without SQI rMSE (bpm)	EM with SQI rMSE (bpm)	EM without SQI rMSE (bpm)
24	1.09 ± 0.87	1.05 ± 0.89	1.09 ± 0.88	1.05 ± 0.89	1.09 ± 0.87	1.05 ± 0.89
18	1.09 ± 0.88	1.05 ± 0.89	1.09 ± 0.90	1.05 ± 0.89	1.09 ± 0.88	1.05 ± 0.89
12	1.61 ± 1.43	1.75 ± 1.64	1.14 ± 0.95	1.05 ± 0.89	1.61 ± 1.43	5.19 ± 5.07
6	2.13 ± 1.81	16.85 ± 14.81	1.26 ± 1.04	1.05 ± 0.89	3.82 ± 3.52	27.72 ± 24.01
0	2.31 ± 1.95	40.80 ± 26.73	1.41 ± 1.17	3.06 ± 2.97	2.92 ± 2.56	50.48 ± 30.35
−6	2.40 ± 2.01	50.23 ± 26.70	2.52 ± 2.32	15.55 ± 14.79	2.64 ± 2.27	53.40 ± 24.29

only one of the 1444 true alarms and was estimated correctly in 281 of the 433 episodes of false alarms. That is, over 99.9% of the true tachycardic episodes were tracked correctly, and 35% of the false episodes were incorrectly tracked as significant heart rate increases.

4. Discussion

SQI plays an important role in eliminating the effect of noise and artifact from HR estimation. When an appropriate signal quality threshold, SQI_{th}, is selected, noisy data can be eliminated from the estimate and robust tracking of the HR variations can be performed. Table 2 illustrates the results of a comparison between the Kalman filtered HR estimation with and without SQI control. It is evident that when there is no SQI control, the error is much higher when the SQI is low (and the noise level is high).

The result of multi-lead ECG analysis is usually better than single-lead analysis (Kaiser and Findeis 2000). This is true when there is only a low level of noise present (when the SNR is 24, 18 and 12) and for some noise types, such as for the *MA* noise or the *EM* noise (as shown in table 1). If the noise level is high, such as for an SNR of 6, 0 and −6 for *EM* noise, each channel is corrupted by serious noise and additional noisy leads cannot provide more valuable information. In this case, multi-lead analysis (SHm) is not better than single-lead analysis (SH1). However, the Kalman filter still provides a superior result in all scenarios, since it rejects noisy segments.

Data fusion of the noisy ECG with the clean ABP can therefore provide a much improved estimation of HR even when the ECG is corrupted completely by noise and artifact. The residual of Kalman filter and the SQI estimates provide reliable criteria for the data fusion stage. Although the *ECGSQI* metric ranges continuously between 0 and 1, there are two interesting discontinuities in the metric. First, $ECGSQI = \eta$ denotes a boundary between extremely high quality ECG (perhaps useful for diagnostic purposes and beat classification) and moderate quality ECG, which can still be used for HR estimation (and maybe even ventricular rhythm analysis). However, if the *ECGSQI* drops too low, the ECG becomes untrustworthy, even for heart rate estimation. This second threshold turns out to be around $ECGSQI = 0.5$. Therefore, the *ECGSQI* metric allows for three levels of ECG quality, which can be used for three different purposes.

Robust and accurate HR estimation has also been demonstrated at both extremely low and extremely high heart rates by considering episodes of bradycardia and tachycardia. However, the performance is not symmetric and the data fusion method presented here tracks episodes of tachycardia more accurately than bradycardia, since the use of a pulsatile waveform almost always allows one to discard the erroneously high ECG-derived heart rates. Conversely, the HR fusion method is fooled by artifact into tracking low heart rates less frequently than for high

heart rates. This is probably due to the higher preponderance of artifacts during tachycardic rather than bradycardic episodes. This algorithm could therefore serve as a first stage filter to false alarm reduction algorithms in the ICU. The *kSQI* metric will not differentiate sinusoidal-like arrhythmias (such as ventricular flutter), from certain noises, both of which can manifest low *kSQI* values (generally below 2). However, since all of the *ECG SQI* metrics have low SQIs during such arrhythmias (except perhaps a frequency-adjusted *sSQI*), they should not be used to reject such alarms. In such cases, reference must be made to a correlated pressure waveform (such as the ABP) to determine the truth of such an alarm (Clifford *et al* 2006, Aboukhalil *et al* 2007).

We should note at this point that some of the thresholds for the individual signal quality metrics have not been tuned to produce optimal results. Further tests on more types of data may yield a more accurate method. In particular, the *ECGSQI* metric is undersensitive to electrode motion (since it is QRS-like in morphology), see figure 5. Our technique would therefore benefit from isolating a metric (or inventing a new metric) that is particularly sensitive to electrode motion, such as the deviation of the QRS morphology away from an average template. However, the current incarnation of the algorithm is still noise-type agnostic for moderate to high SQI values ($ECGSQI \geqslant 0.7$).

It should also be noted that in our approach the KF uses a trivial model of the cardiovascular system (the heart rate at each epoch is approximately the same as the next epoch). It is possible that more complicated models of the HR and ABP could be used and, in particular, the model could be extended to treat the HR and ABP simultaneously, with a four-dimensional state: HR, systolic blood pressure (SBP), mean blood pressure (MBP) and diastolic blood pressure (DBP). Known relationships between these parameters may be utilized to improve the estimate. Simple models of their relationship to cardiac output could also be used.

5. Conclusion

We have developed a robust HR estimation method based on ECG and ABP beat detection, signal quality analysis, Kalman filtering and data fusion. By applying this method to an extensive database of ICU signals, we have demonstrated that good HR estimation is possible (rMSE $\leqslant 1 \pm 0.9$ bpm), even in the presence of high levels of noise and artifact, and during episodes of extreme bradycardia and tachycardia. By calibrating our SQI output on known data and noise, it is possible to estimate the error in a given HR estimate on new ECG data, by determining the SQI value with our metrics. These robust SQI metrics therefore provide a method for evaluating the physiological signal quality and can be used as guidance for identifying periods of data for estimating derived health metrics, data mining activities, inputs to cardiovascular models, selecting training data in machine learning tasks and false alarm reduction in ICU monitors. Furthermore, the general framework proposed in this paper is applicable to any multi-parameter input time series where SQIs can be estimated, such as BP evaluation, cardiac output estimation or even non-biological parameter tracking.

Acknowledgments

This work was supported by the US National Institute of Biomedical Imaging and Bioengineering (NIBIB) under grant number R01 EB001659 and the Information & Communications University, Daejon, South Korea, under grant number 6914565. The content of this paper is solely the responsibility of the authors and does not necessarily represent the

official views of the NIBIB or the NIH. The authors would like to acknowledge Drs Tarassenko, Townsend, Sun and Zong, whose work contributed significantly to this research.

References

Aboukhalil A, Nielsen L, Saeed M, Mark R G and Clifford G D 2007 Reducing false alarm rates for critical Arrhythmias: a large-scale MIMIC II study *J. Biomed. Inform.* Special issue on computerized decision support for critical and emergency care, at press

Allen J and Murray A 1996 Assessing ECG signal quality on a coronary care unit *Physiol. Meas.* **17** 249–58

ANSI/AAMIEC57 1998 *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms* American National Standards Institute

Brown R G 1983 *Introduction to Random Signal Analysis and Kalman Filtering* (New York: Wiley)

Chambrin M C 2001 Alarms in the intensive care unit: how can the number of false alarms be reduced? *Crit. Care* **5** 184–8

Chen L, McKenna T, Reisner A and Reifman J 2006 Algorithms to qualify respiratory data collected during the transport of trauma patients *Physiol. Meas.* **27** 797–816

Clifford G D 2006 Linear filtering methods *Advanced Methods and Tools for ECG Data Analysis* ed G D Clifford, F Azuaje and P E McSharry (Norwood, MA: Artech House) chapter 5

Clifford G D, Aboukhalil A and Mark R G 2006 Using the blood pressure waveform to reduce critical false ECG alarms *Comput. Cardiol.* **33** 829–32

Clifford G D and Li Q 2007 What are heart rate and blood pressure, and when are errors significant? *MIT Technical Report*

Clifford G D, McSharry P E and Tarassenko L 2002 Characterizing abnormal beats in the normal human 24-hour RR tachogram to aid identification and artificial replication of circadian variations in human beat to beat heart rate *Comput. Cardiol.* **29** 129–32

Ebrahim M H, Feldman J M and Bar-Kana I 1997 A robust sensor fusion method for heart rate estimation *J. Clin. Monit.* **13** 385–93

Feldman J M, Ebrahim M H and Bar-Kana I 1997 Robust sensor fusion improves heart rate estimation: clinical evaluation *J. Clin. Monit.* **13** 379–84

Friesen G M, Jannett T C, Jadallah M A, Yates S L, Quint S R and Nagle H T 1990 A comparison of the noise sensitivity of nine QRS detection algorithms *IEEE Trans. Biomed. Eng.* **37** 85–98

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** e215–20

Hamilton P S and Tompkins W J 1986 Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database *IEEE Trans. Biomed. Eng.* **33** 1157–65

He T, Clifford G D and Tarassenko L 2006 Application of independent component analysis in removing artefacts from the electrocardiogram *Neural Comput. Appl.* **15** 105–16

Jakob S, Korhonen I, Ruokonen E, Virtanen T, Kogan A and Takala J 2000 Detection of artifacts in monitored trends in intensive care *Comput. Methods Programs Biomed.* **63** 203–9

Kaiser W and Findeis M 2000 Novel signal processing methods for exercise ECG *Int. J. Bioelectron.* **2** 61–65

Kohler B U, Hennig C and Orglmeister R 2002 The principles of software QRS detection *IEEE Eng. Med. Biol. Mag.* **21** 42–57

Lawless S T 1994 Crying wolf: false alarms in a pediatric ICU *Crit. Care Med.* **22** 981–5

Li Q, Mark R G and Clifford G D 2007 Realistic artificial arterial blood pressure artifact algorithms and an evaluation of a robust blood pressure estimator, in preparation

Mendelson Y 1992 Pulse oximetry: theory and applications for noninvasive monitoring *Clin. Chem.* **38** 1601–7

Moody G B and Mark R G 1997 Integration of real-time and off-line clinical data in the MIMIC database *Comput. Cardiol.* **24** 585–8

Moody G B and Mark R G 2001 The impact of the MIT–BIH arrythmia database *IEEE Eng. Med. Biol. Mag.* **20** (3) 45–50

Moody G B, Muldrow W E and Mark R G 1984 A noise stress test for arrhythmia detectors *Comput. Cardiol.* **11** 381–4

Murthy V K, Grove T M, Harvey G A and Haywood L J 1978 Clinical usefulness of ECG frequency spectrum analysis *Computer Application in Medical Care: 1978 Proceedings* pp 610–2

Oefinger M B 2006 Monitoring transient repolarization segment morphology deviations in Mouse ECG *PhD Thesis* Massachusetts Institute of Technology, Cambridge, MA

Saeed M, Lieu C, Raber G and Mark R G 2002 MIMIC II: a massive temporal ICU patient database to support research in intelligent patient monitoring *Comput. Cardiol.* **29** 641–4

Sittig D F and Factor M 1990 Physiologic trend detection and artifact rejection: a parallel implementation of a multi-state Kalman filtering algorithm *Comput. Methods Programs Biomed.* **31** 1–10

Sun J X, Reisner A T and Mark R G 2006 A signal abnormality index for arterial blood pressure waveforms *Comput. Cardiol.* **33** 13–6

Tarassenko L, Mason L and Townsend N 2002 Multi-sensor fusion for robust computation of breathing rate *Electron. Lett.* **38** 1314–16

Tarassenko L, Townsend N, Clifford G, Mason L, Burton J and Price J 2001 Medical signal processing using the software monitor *Intelligent Sensor Processing* (Ref. No. 2001/050) pp 3/1–3/4

Tarassenko L, Townsend N and Price J D 2003 Combining measurements from different sensors *United States Patent* Application Number 311250 (http://www.freepatentsonline.com/20030187337.html)

Townsend N 2001 Merging multiple estimates to form a single estimate *Unpublished Technical Report* (Oxford: Department of Engineering Science, University Oxford)

Tsien C L and Fackler J C 1997 Poor prognosis for existing monitors in the intensive care unit *Crit. Care Med.* **25** 614–9

Tsien C L, Kohans I S and Mclntosh N 2001 Building ICU artifact detection models with more data in less time *Proc. AMIA Symp.* pp 706–10

Welch G and Bishop G 2004 An introduction to the Kalman filter *Technical Report* Dept. Comp. Sci., University of North Carolina at Chapel Hill, TR95–041

Zong W, Heldt T, Moody G B and Mark R G 2003a An open-source algorithm to detect onset of arterial blood pressure pulses *Comput. Cardiol.* **30** 259–62

Zong W, Moody G B and Jiang D 2003b A robust open-source algorithm to detect onset and duration of QRS complexes *Comput. Cardiol.* **30** 737–40

Zong W, Moody G B and Mark R G 2004 Reduction of false arterial blood pressure alarms using signal quality assessment and relationships between the electrocardiogram and arterial blood pressure *Med. Biol. Eng. Comput.* **42** 698–706