Physiol. Meas. 29 (2008) 81–94 doi:10.1088/0967-3334/29/1/006

# A fuzzy logic algorithm to assign confidence levels to heart and respiratory rate time series

**J Liu[1], T M McKenna[1], A Gribok[1], B A Beidleman[2], W J Tharion[2] and J Reifman[1]**

[1] Bioinformatics Cell, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Materiel Command, Fort Detrick, MD 21702, USA
[2] US Army Research Institute of Environmental Medicine, Natick, MA 01760, USA

E-mail: jeanliulv@yahoo.com, thomas.mckenna2@us.army.mil, agribok@bioanalysis.org, beth.beidleman@us.army.mil, william.tharion@us.army.mil and jaques.reifman@us.army.mil

Received 31 July 2007, accepted for publication 20 November 2007
Published 3 January 2008
Online at stacks.iop.org/PM/29/81

**Abstract**

We have developed a fuzzy logic-based algorithm to qualify the reliability of heart rate (HR) and respiratory rate (RR) vital-sign time-series data by assigning a confidence level to the data points while they are measured as a continuous data stream. The algorithm's membership functions are derived from physiology-based performance limits and mass-assignment-based data-driven characteristics of the signals. The assigned confidence levels are based on the reliability of each HR and RR measurement as well as the relationship between them. The algorithm was tested on HR and RR data collected from subjects undertaking a range of physical activities, and it showed acceptable performance in detecting four types of faults that result in low-confidence data points (receiver operating characteristic areas under the curve ranged from 0.67 (SD 0.04) to 0.83 (SD 0.03), mean and standard deviation (SD) over all faults). The algorithm is sensitive to noise in the raw HR and RR data and will flag many data points as low confidence if the data are noisy; prior processing of the data to reduce noise allows identification of only the most substantial faults. Depending on how HR and RR data are processed, the algorithm can be applied as a tool to evaluate sensor performance or to qualify HR and RR time-series data in terms of their reliability before use in automated decision-assist systems.

Keywords: heart rate, respiratory rate, fuzzy logic, confidence levels, data qualification, sensor validation

  

## 1. Introduction

Advances in sensor and computer technology allow the continuous measurement of physiology variables of active individuals in the field (i.e., not in a laboratory or clinical environment). The objective of such monitoring is to determine the physiologic state of the individual, which can provide benefit in terms of warning the person, or other parties, of present or impending physiologic stress in response to environmental or traumatic injury.

Various circumstances influence the measurement of physiological variables for field applications. The sensors and processing unit carried by an individual must be small. Individuals must be able to wear the unit continuously without having to carry heavy batteries or replace them frequently; therefore, power consumption by the unit must be low. It is usually not possible to collect redundant measures of each physiology variable or to collect a wide range of physiology variables at one time. Furthermore, the data are susceptible to motion artifacts resulting from movement by the subject or sensor and to hardware faults, such as intermittent signals from damaged sensors or leads. The net effect of these constraints is that field-collected data may be sparse compared with the amount that can be collected in the laboratory and may be of questionable reliability for making medical decisions.

The United States Army Research Institute of Environmental Medicine (USARIEM, Natick, MA) is developing a wearable suite of sensors, mostly physiological, and a data processing unit that is, in total, termed the Warfighter Physiological Status Monitoring (WPSM) system. The WPSM system, in its current configuration, can monitor heart rate (HR), respiratory rate (RR), skin temperature, body position and motion, and can detect a ballistic impact to the body, such as might occur from a bullet. The WPSM system is expected to perform several roles in the management of health care on or off the battlefield. These would include (1) aiding in the prevention of injuries, (2) determining the live/dead status of the soldier and (3) if an injured soldier is alive, the system should send information to a medic to help facilitate medical treatment of that soldier (Hoyt *et al* 2002). Therefore, an important functionality of the WPSM system is to operate as a vital-sign monitor.

The realities that field-collected data are likely to be sparse and noisy, while at the same time medical decisions will hinge on that data, require that a certain level of confidence must exist in their quality before they are used to diagnose or predict the physiologic state of an individual. A significant number of quality assessment and artifact detection algorithms have been proposed in the literature. For example, HR and blood-pressure data were assessed (Cao *et al* 1999) using three different types of artifact detectors: limit-based, deviation-based and correlation-based. The algorithm produced high sensitivity and specificity (over 90%) for both HR and blood pressure artifacts; however, the system was developed and tested on data collected from preterm infants, so motion artifacts were not a significant concern. In general, most quality-assessment algorithms require the availability of the underlying waveforms (electrocardiogram (ECG) or respiratory) to qualify the reliability of the derived vital-sign data. For example, fuzzy logic has been applied to monitor the quality of vital-sign data by integrating ECG waveform, oxygen partial pressure and pulse oximeter data using fuzzy rules (Wolf *et al* 1996). Many other approaches have reported high (over 90%) sensitivity and specificity of vital-sign data qualification based on waveform information (Xu and Schuckers 2001, Jiang *et al* 2007). However, the availability of waveform data cannot be assured in unstructured field applications. Another obstacle for field deployment of waveform-based algorithms relates to the computational limitations of wearable devices. For example, the method to detect artifacts reported by Park *et al* (2002) requires a multi-step optimization of the ECG histograms, which is computationally intensive and cannot be performed by existing wearable devices.

We have developed a physiology-based, fuzzy logic algorithm to assign a confidence level to HR and RR time-series data as they are collected, with the assumption that neither ECG nor respiratory waveforms are available. The algorithm may be applied to either raw or filtered vital-sign data, depending on whether the measured data point value and its associated confidence level are required or whether it is preferable to subject the data to signal processing procedures before determining a derived data point and its associated confidence level.

## 2. Datasets

The HR and RR time-series data used in this study were collected at the USARIEM (Beidleman *et al* 2004). Eight non-smoking volunteers [21 years (SD 3) age, 76 kg (SD 9) weight, 175 cm (SD 5) height, mean and standard deviation (SD)] participated as subjects in this study. The subjects wore four sensors concurrently to collect HR and RR for approximately 4 h a day while they engaged in low- (sit, lie, stand), medium- (walk, sit-ups, push-ups, jumping jacks) and high-level (run) activities. Two of the sensors, incorporated into a VivoMetrics Lifeshirt (Ventura, CA), measured HR and RR, and two different sensors provided simultaneous, redundant measures of HR (Schiller Cardiovit AT-6 ECG machine; Schiller Inc., Baar, Switzerland) and RR (SensorMedics Model 2900 metabolic cart; SensorMedics, Yorba Linda, CA). The original objective for the data collection was to test the reliability and validity of the HR and RR measures by the VivoMetrics Lifeshirt, which incorporates the sensors in a wearable garment. The Schiller Cardiovit AT-6 and the SensorMedics metabolic cart are standard laboratory devices for the collection of physiological measurements and were used to set the parameters of the algorithm. A new data record was generated at every change in HR or RR detected by any of the systems, resulting in HR and RR sampling rates from 1 to 4 s, with an average rate of around 2 s. This sampling protocol results in essentially 'instantaneous' measures of the variables, which effectively unmask distinct, transitory faults that are characterized as measures that vary from true, reasonable values.

Heart rate from the VivoMetrics system was obtained by using three ECG electrodes positioned on the chest just above the left and right nipples and on the side of the left abdomen. Respiration from the VivoMetrics system was obtained through respiratory inductive plethysmography that uses changes in volume of the cross-sectional area of the rib cage and abdomen. These measures are obtained by thin insulated wires embedded in the elastic bands woven into the VivoMetrics system. Low-voltage electrical current is passed through the wire, creating an oscillating circuit. In response to respiratory movements, the electrical sensors generate different magnetic fields that are converted into proportional voltage changes and, through proprietary algorithms from VivoMetrics, a conversion into RR is determined. The SensorMedics metabolic cart recorded RR every time a breath was taken by measuring inspired air through a mouthpiece with the nose clipped off. The SensorMedics cart registered the minute-by-minute RR and associated respiration waveform. Use of the SensorMedics cart to assess RR has previously proven to be reliable and valid (Macfarlane 2001, Unnithan *et al* 1994). Heart rates obtained from the Schiller used standard three-lead ECG, which were placed next to the ECG electrodes from the VivoMetrics system. The Schiller machine meets ECG instrument specifications of the American Heart Association (Bailey *et al* 1990). The VivoMetrics and Schiller systems provide ECG waveforms, and the VivoMetrics and SensorMedics systems provide respiration waveforms. Heart rate for both systems was determined from the ECG. The ECG and respiration waveforms were displayed and examined for any abnormalities (either for possible volunteer health issues associated with the testing or possible equipment malfunctions) during testing. However, due to hardware storage limitations, ECG and respiration waveforms were not saved.

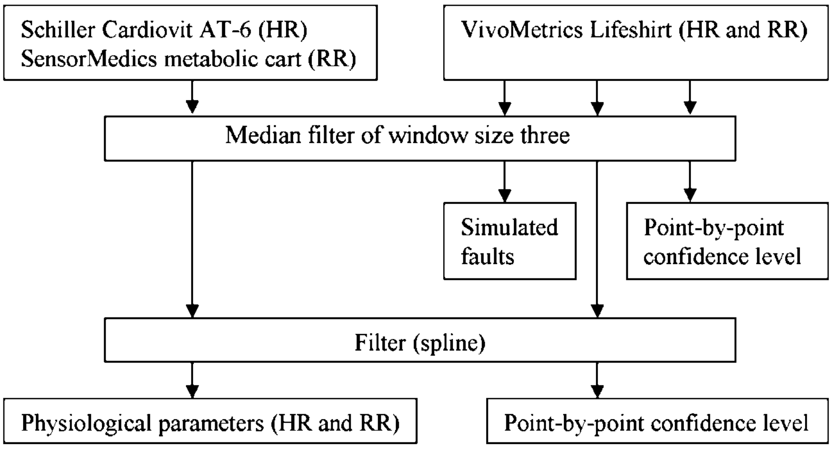


**Figure 1.** Datasets and processes used to develop and evaluate a fuzzy logic algorithm to assign confidence values to HR and RR.

The simultaneous, but separately acquired, measures of HR and RR from each subject resulted in three datasets, named after the systems with which they were collected: VivoMetrics, Schiller and SensorMedics (figure 1). All of the datasets were filtered with a median filter of window size three to remove single data point outliers and then resampled to a 2 s sampling rate to compensate for differences in sensor sampling frequencies. The Shiller and SensorMedics datasets were further filtered with a cubic spline smoothing filter, which is a standard signal processing method to reduce noise in a noisy dataset (Wahba 1990). These filtered datasets were used to calculate physiological parameters to construct the membership functions of a fuzzy logic algorithm that assigns confidence values to HR and RR data points. In contrast, the VivoMetrics dataset was used as a test bed to evaluate the ability of the fuzzy logic algorithm to identify actual and simulated low-confidence data points, reflecting unreliable measurements. The VivoMetrics dataset was subsequently filtered with the cubic spline smoothing filter, and the ability of the algorithm to identify low-confidence data points was retested using the smoothed dataset. In effect, the Schiller and SensorMedics datasets were used to develop the fuzzy logic algorithm, while the VivoMetrics dataset was used to evaluate the algorithm.

## 3. Fuzzy logic estimation of data point confidence level

### *3.1. Fuzzy logic structure*

In the fuzzy logic-based algorithm, five block-processing elements capture (1) the relationships between HR and RR, (2) the quality of the measures for HR and RR and (3) the resulting confidence for the HR and RR values (figure 2). The top of the figure indicates that the relationships between HR and RR are evaluated; a true relationship indicates that the HR and RR have a physiologically reasonable ratio to each other and that they also have similar trend directionality. The bottom of the figure indicates that the reliability of the HR and RR measures is evaluated; a true measure means that a HR or a RR measure is a reliable reading from a sensor. The HR and RR measures include the mean or median, rate of change (i.e., slope), noise in the signal and whether the signal changes over time. The final membership outputs

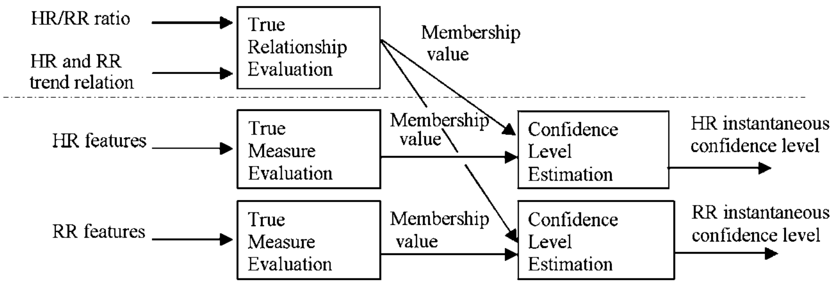


**Figure 2.** Block structure of the fuzzy logic algorithm to estimate confidence levels. The dotted line separates the parts of the algorithm that evaluate the relationships between the HR and RR (above the line) and the quality of the HR and RR measurements (below the line).

from the confidence level estimation blocks (right side of figure 2) represent the likelihood of a HR or a RR value indicating a true physiological condition, which we define as the confidence level. Because the confidence levels are based on short 15 s windows, they are termed instantaneous confidence levels. The 15 s window represents the current requirement for the WPSM system for reporting the status of a soldier.

### *3.2. Input data features*

All data processing and analysis procedures were performed sequentially in 15 s long windows. The fuzzy logic algorithm requires a total of ten input features. Two of the features represent the relationships between HR and RR; these are ratios and trends. The remaining eight features are derived from measures of HR and RR.

*Ratio*. The HR/RR ratio captures the relative coincidence between HR and RR, when both fall within a physiologically reasonable range. When the measured HR and RR establish an unreasonable relationship to each other, although neither one is obviously wrong, they are deemed unreliable and a low membership value is assigned to the true relationship. Alternatively, if either measure is apparently wrong (out of a conservative range of normal physiological values), then the HR/RR ratio is set to a default value of 4, which disables the ratio evaluation, and only the true measure evaluation of each variable (instead of considering the relationship evaluation) will determine the final confidence level. The HR/RR ratio is calculated as

$$\text{HR/RR ratio} = \begin{cases} \dfrac{H}{R}; & \dfrac{45 \leqslant H \leqslant 190}{10 \leqslant R \leqslant 70} \\ 4; & \text{otherwise} \end{cases} \tag{1}$$

where $H$ and $R$ are 15 s values for mean HR and median RR, respectively.

*Trend.* In general, it is expected that directional changes of HR and RR are correlated, taking into account time lags and a certain degree of individual manipulation of RR (e.g. ‘pacing’ during exercise). If HR and RR trends are opposed, a low membership value is assigned to the true relationship. This feature is based on 1 min slopes for HR and RR. The HR slope is estimated by a least-squares error (LSE) regression on data points in the current 1 min window. The RR slope is calculated by taking the median RR in the current 15 s window and

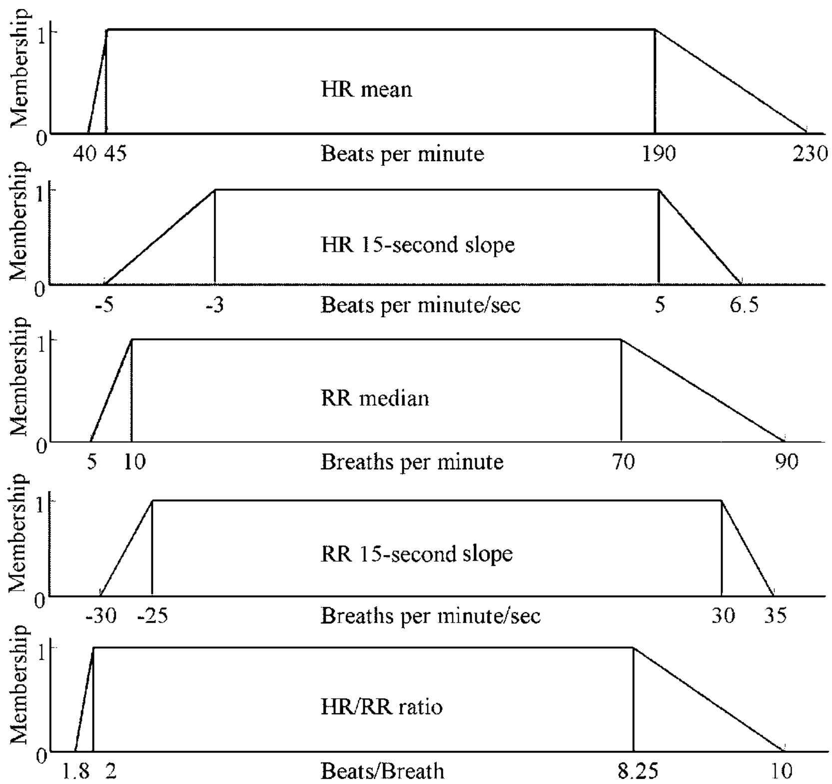


**Figure 3.** Membership functions for mean HR, HR 15 s slope, RR median value, RR 15 s slope and HR/RR ratio.

subtracting it from the median RR in the 15 s window at 60–45 s before the current time, and dividing by 45 s.

*Measures.* The following four features are extracted from each HR and RR time series: (1) mean (or median for RR), calculated from sequential 15 s windows, (2) 15 s slope, which is calculated by LSE regression within a 15 s window, (3) noise, calculated by obtaining the residuals (i.e., the difference between HR/RR measurements and their regression over 15 s windows), and by computing the variance of the residuals assuming the mean is zero, and (4) a constant signal interval, which detects unchanging HR/RR measures; it is a feature to determine whether a sensor has failed and is stuck at the same value.

### *3.3. Membership function design*

We employed two approaches to construct the fuzzy logic membership functions. Some features, for example, the HR and the RR mean, median, or slopes, have physiology-based upper and lower limits. The membership functions for these features were defined based on these limits; data inside this range are considered reasonable with a degree of 1, while data outside the range are considered reasonable with a decreasing degree, as they get farther away from the cut-off range (figure 3). We employ trapezoidal membership functions for this type of features. We believe that the trapezoidal function is an appropriate and convenient approximation to describe the fuzziness attributed to physiologic variables, since it assigns a

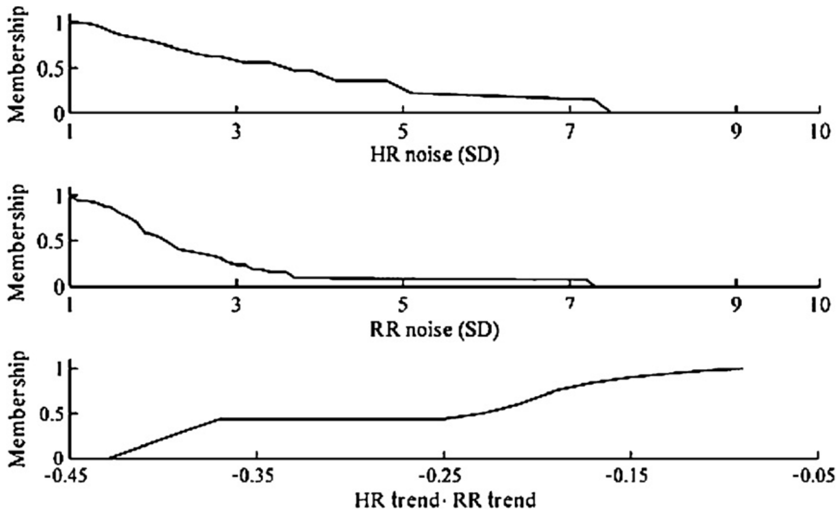


**Figure 4.** Data-driven membership functions for HR noise, RR noise and HR trend in beats per min/s (BPM $s^{-1}$) × RR trend in breaths per min/s (BrPM $s^{-1}$).

possibility of 1 to the normal range of physiologic values, and a gradually decreasing degree of membership to the values outside that range.

The physiological limits were defined using the Schiller and SensorMedics datasets, which were filtered with a median filter of window size 3, resampled to a 2 s sampling rate, and then filtered with a cubic spline smoothing filter. The regularization parameter used in the cubic spline filter was selected by cross-validation (Wahba 1990). This filter removes noise to yield smoothly changing estimates of the HR and RR that approximate the true values for these variables. The physiological limits extracted from the datasets include the HR mean, and RR median, slopes, and the ratio of mean HR to median RR (figure 3). For example, the slope values for both HR and RR were derived from instantaneous derivatives obtained on smooth data. The remaining limits were identified by visual determination of the maximum and minimum physiologically possible means, medians and ratios of mean HR to median RR. The minimum and maximum values were set as limits for full membership (i.e., degree of 1); the partial membership limits (i.e., degree less than 1) were subjectively set after review of the literature and examination of the raw data. These membership functions can be considered as physiology based, since they are generally sensor independent.

Other features are strongly affected by factors such as sensor quality, sampling rate or motion-induced recording artifacts, rather than physiological limits. In the simplest case, the membership function for a constant signal interval feature is defined as a linear decrease after 30 s of constant signal, with zero membership after 60 s. The membership functions for HR noise, RR noise, and the HR and RR trend relationship (figure 4) are derived from their distributions through a transformation based on mass assignment theory (Shanahan 2000). Mass assignment, a set-based probability function, builds a bridge between a probability density function and a fuzzy set membership function. The noise estimate used in the probability density function to develop the noise membership function was generated by first estimating the amount of noise in the raw data with the spline smoothing technique to produce a filtered signal and then by determining the variance of the residuals between the filtered signal and the raw data. The spline regularization parameter, which controls the degree of smoothing, was selected using the cross-validation method (Wahba 1990). This method effectively trades off the squared bias and the variance of the filtered signal. We also checked the residuals for whiteness to ensure that the filtered signal was neither under- or oversmoothed.

### *3.4. Fuzzy rules*

Five fuzzy rules correspond to the blocks in figure 2. One rule evaluates the HR and RR relationships, and two rules evaluate the quality of the HR and RR measurements. Two additional rules estimate the confidence levels for HR and RR. The rules operate on HR and RR measurements using a logical '*AND*' operator to produce the final confidence for HR and RR values. The rules are

(1) *IF* the HR/RR ratio is reasonable *AND* the HR and RR trend relation is reasonable, *THEN* the relationship is true.
(2) *IF* the HR mean value is reasonable *AND* the HR 15 s slope is reasonable *AND* the HR noise is reasonable *AND* the HR constant signal interval is reasonable, *THEN* the measure for HR is true.
(3) *IF* the RR median value is reasonable *AND* the RR 15 s slope is reasonable *AND* the RR noise is reasonable *AND* the RR constant signal interval is reasonable, *THEN* the measure for RR is true.
(4) *IF* the relationship is true *AND* the measure for HR is true, *THEN* the confidence for HR is true.
(5) *IF* the relationship is true *AND* the measure for RR is true, *THEN* the confidence for RR is true.

The membership function for any evaluation being true is a constant value of one, corresponding to a Sugeno-type fuzzy inference (Sugeno 1985). The output levels for the first three rules are weighted by the firing strength of the rules as determined by the membership functions for inputs to the rules. In this fuzzy logic model, the logical *AND* operator performs as a minimum operation for all rules. The membership value for the confidence level is assigned to the corresponding HR or RR variable every 15 s, providing an instantaneous confidence level.

## 4. Analysis of algorithm performance by receiver operating characteristic (ROC) curves

### *4.1. Simulated faults*

To validate the algorithm, four types of simulated faults were superimposed, individually and in combination, on the median filtered and resampled VivoMetrics dataset. When the faults were superimposed individually, 100 faults were superimposed on the data from each subject; when superimposed in combination, 25 faults of each type were superimposed on the data from each subject. The superimposed faults yielded a fault rate of approximately 20% of the data points for each subject, for both HR and RR. The magnitudes of the faults were selected to moderately exceed normal physiological limits.

(1) *Spikes with fixed amplitude*. The spikes are two data points in duration in 15 s windows, with amplitudes based on the SD of HR and RR noise in the Schiller and SensorMedics datasets. The maximum SD are 7.2 beats per minute (BPM) and 7.5 breaths per minute (BrPM) for HR and RR, respectively. Their corresponding 95% confidence limits, $\pm 15$ BPM and $\pm 15$ BrPM, respectively, were selected as the amplitude of the spikes. The spikes were randomly superimposed in random positive and negative orientation onto the HR and RR datasets.

(2) *Random noise with zero mean and a preset SD*. Noise sampled from normal distributions with fixed SD of 15 BPM and 15 BrPM for HR and RR, respectively, was superimposed into randomly selected 15 s windows in the datasets.

(3) *Abnormal slopes*. The maximum normal acceleration or deceleration of HR and RR, based on derivatives from the Schiller and SensorMedics datasets over the 15 s windows, was 2.6 BPM $s^{-1}$ for HR and 1.0 BrPM $s^{-1}$ for RR, respectively. The simulated abnormal slope faults were set at twice these maximum values or 5.2 BPM $s^{-1}$ and 2.0 BrPM $s^{-1}$. Abnormal slopes, 15 s long with random positive or negative direction, were inserted into the test datasets at random locations.

(4) *Contradictory trends between HR and RR*. Pairs of contradictory trends (slopes), 1 min in duration and in opposite direction, were randomly inserted into the HR and RR time series at the same time points. The HR slopes were 1.1 BPM $s^{-1}$ and the RR slopes were 0.4 BrPM $s^{-1}$, which are physiologically normal rates.

### *4.2. ROC curves*

The ability of the algorithm to detect the superimposed faults was quantified by ROC curves (Obuchowski 2003). Ideally, faults would be superimposed onto a fault-free dataset and the detection performance of the algorithm assessed. However, the datasets are not fault free, and no method is available to provide objective, *a priori* labeled faults without additional information provided by either ECG waveform or respiratory waveform. Therefore, the ROC curves were constructed by comparing the confidence values assigned by the algorithm to data points altered by the superimposed faults with the original, unaltered data point confidence values, using a set of thresholds ranging from −0.10 to 1.00, with increments of 0.01. In this application, changes from original data point confidence levels indicate faults. The area under the ROC curve (AUC) was calculated by trapezoidal integration to summarize detection performance with a single score. The ROCs were constructed for each subject based on 100 replicates (i.e., 100 faults were randomly inserted in a subjects data, the AUC determined, and the process repeated a total of 100 times); next, the AUCs were averaged over all subjects to obtain the algorithm performance for a specified fault.

## 5. Results

### *5.1. Fault detection*

The 'instantaneous' property of the measures of HR and RR in the VivoMetrics dataset results in the detection of a large number of pre-existing low-confidence data points (i.e., faults) by the algorithm before the simulated faults are superimposed (figure 5). Furthermore, the superimposed faults may occasionally correct pre-existing faults (e.g., a superimposed upward-directed spike will correct a pre-existing downward-directed spike). Under these constraints, the detection of spikes and abnormal slopes was acceptable, whereas that for random noise and contradictory trend faults was not (table 1). Because the SD of the Gaussian distribution used to simulate random noise was set at twice the maximum physiologic SD, 68% of the superimposed noise had an amplitude of less than 1 SD, which is a property similar to true physiological values, making it difficult to discriminate noise from true data. Similarly, the moderate detection performance for contradictory trends was likely due to superimposing a relatively low-slope trend onto the noisy data. The uniformly lower fault detection in RR versus HR data is also likely due to noise in the signal; the signal-to-noise ratio for RR data

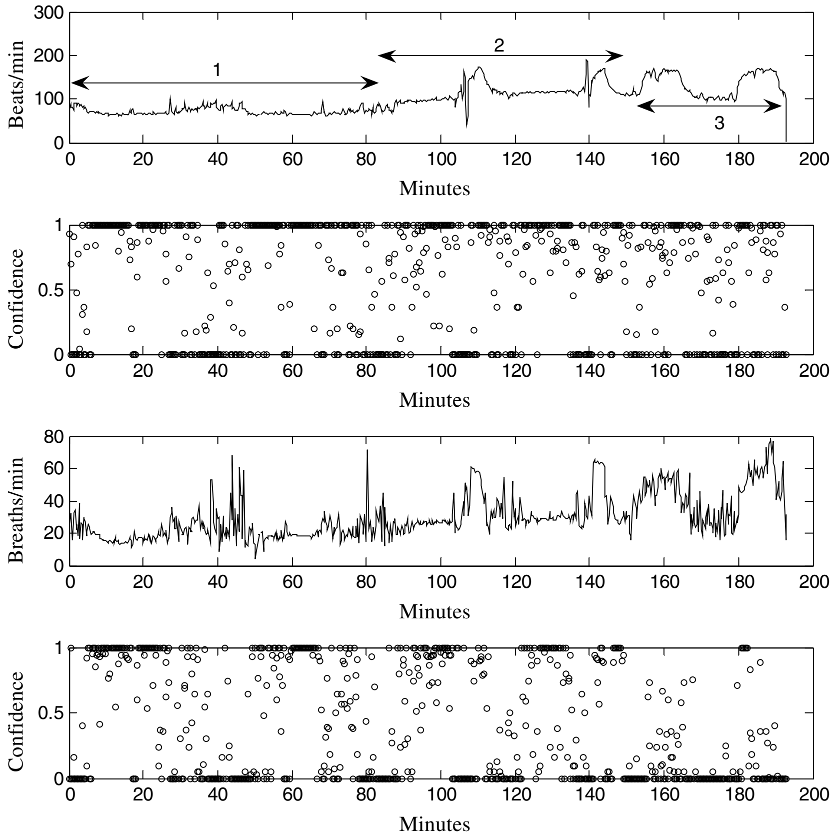


**Figure 5.** The top two panels show the HR and corresponding confidence levels for a subject engaged in low-level activities (1: lie and sit on cot, stand, take a break and then repeat the exercises), medium-level activities (2: walk on treadmill, sit-ups, push-ups, jumping jacks, take a break and then repeat exercises) and high-level activity (3: run on treadmill, take a break and repeat). The data were collected by the VivoMetrics system and were median filtered and resampled to 2 s intervals. The bottom two panels show the RR rate of the same subject and associated confidence levels.

**Table 1.** Performance of the algorithm in detecting simulated faults superimposed on the median-filtered VivoMetrics dataset. Detection performance is quantified by the area under the curve (AUC) of receiver operating characteristic curves and is expressed as the mean and SD for 100 replicates per subject, over eight subjects.

| Simulated fault | AUC for faults in HR | AUC for faults in RR |
|---|---|---|
| Spike | 0.83 (SD 0.03) | 0.76 (SD 0.05) |
| Random noise | 0.75 (SD 0.01) | 0.67 (SD 0.04) |
| Abnormal slope | 0.84 (SD 0.01) | 0.80 (SD 0.04) |
| Contradictory trend | 0.72 (SD 0.02) | 0.69 (SD 0.03) |
| All | 0.80 (SD 0.03) | 0.75 (SD 0.05) |

is about half that of HR data (mean of 1.7 (SD 0.4) versus mean of 3.3 (SD 0.7), over all subjects), which will tend to mask the superimposed faults in pre-existing noise.

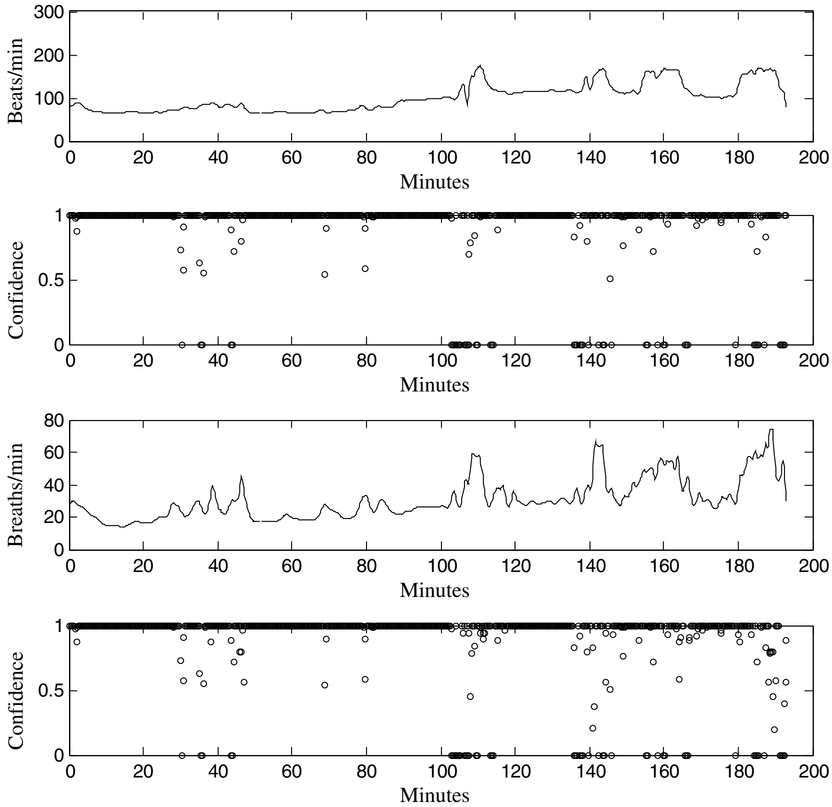


**Figure 6.** Heart rate and the corresponding confidence level (top two panels) along with RR and the corresponding confidence level (bottom two panels) for the same subject as in figure 5. The data were collected by the VivoMetrics system and were median filtered, resampled to 2 s intervals and filtered with a spline smoothing filter before application of the algorithm.

### *5.2. Application of the fuzzy logic algorithm to progressively filtered data*

The algorithm was applied to the VivoMetrics HR data from a representative subject undertaking the full range of physical activities. Confidence levels were determined after the data were processed by median filtering and resampling (figure 5) and after additional filtering with a spline smoothing filter (figure 6). The algorithm is very sensitive to noise in the non-smoothed data (figure 5); if data points above an arbitrary confidence level threshold of 0.5 are taken as reliable, then only 63% of the HR data and 42% of the RR data are acceptable. In contrast, if noise in the data is reduced by filtering the data with a spline filter, then 92% of both HR and RR data are acceptable at the same threshold, and the low-confidence data points are generally associated with activities that are likely to cause motion artifacts (figure 6).

## 6. Discussion

An algorithm to assign confidence levels to physiologic time-series data has two potential applications: the evaluation of sensor performance and the screening of reliable data for use in a downstream decision-assist application. In the first case, the objective is often to assign a confidence value to the state of certain observation, e.g., QRS complex, computed by a physiological monitor based on actual sensor measurements. The challenge is that rates are susceptible to large point-by-point excursions because they are computed over short, independent time windows. However, this property is not an impediment to using the algorithm to evaluate sensor reliability. For instance, a cluster of low-confidence points can be useful for identifying a sensor's susceptibility to motion-induced artifacts when a subject undertakes a particular motion or body orientation. Similarly, a consistent run of high-confidence data points will indicate optimal sensor function. Minimal, if any, processing of the raw data is necessary when the algorithm is used in this capacity. Figure 5 exemplifies the results for this kind of application.

It is more challenging to use an algorithm that generates point-by-point confidence level information to provide reliable data to applications that must make a decision based on the data, such as a decision-assist application. To be useful in these applications, a system that uses physiological data must take into account how much credence can be placed on it while at the same time avoid point-by-point oscillations in the confidence of the output, which can limit the utility of the output for decision-assist purposes. This is necessary because potential short-term, non-critical faults will flag the data or system as unreliable. Too many false alarms degrade a system to a level where it is of little worth (Edworthy and Hellier 2006).

The fuzzy logic algorithm can be employed in two modes to avoid rapid changes in confidence level while still providing useful information for an actionable purpose. If data-point-by-data-point output from the application is not necessary (e.g., mean values over a period of time are acceptable), then only data points with confidence levels above a threshold can be used by the application. In this case, the fuzzy logic algorithm acts as a filter, passing on only reliable data. In the second mode, the data are filtered to reduce noise in the signal, and then confidence levels are assigned by the algorithm. In this instance, point-by-point data and their associated confidence levels are available, at the cost of replacing the original measured data values with those that rely upon the performance characteristics of the filter used, as shown in figure 6. Similarly, the VivoMetrics system accurately estimates respiratory variables during treadmill exercise using values averaged over 1 min (Witt *et al* 2006). In both cases, averaging or spline smoothing acts as a low-pass filter to reduce noise in the signal.

It is likely that methods to identify reliable data will be required before decision-assist applications can routinely be implemented in the field, because it is difficult to consistently acquire accurate physiology waveform signals in such dynamic environments. For instance, during helicopter transport of more than 700 injured patients from the location of injury to a hospital, less than half of the collected ECG and less than 25% of the respiratory waveform data from which HR and RR are calculated, respectively, were evaluated as good quality (Yu *et al* 2006, Chen *et al* 2006).

There are advantages of applying a fuzzy logic algorithm to calculate the confidence placed on data points measured by physiology monitoring systems. It is possible to formalize and simulate the domain knowledge of those skilled in a medical discipline to construct the membership functions, and the method can efficiently take into account several variables and perform 'weighted merging' of differing influence of the variables. This process can yield an algorithm that captures the nonexplicit nature of clinical decision making (Bates and Young 2003). Because the membership functions are derived empirically, fault detection specificity can be increased, if desired, by abridging the membership function span. Changes in the membership functions can also be used to tailor the algorithm to specific groups of subjects (e.g., sedentary versus athletic). Other advantages include the fact that the Sugeno system notation is very compact and efficient, and the simple computation and evaluation of features and membership functions make the method appropriate for computational resources likely to be encountered in field applications.

The work presented here has methodological and technological limitations. The main methodological limitation relates to the fact that the 'quality' of the original test data was not known. The unavailability of ECG and respiratory waveform recordings precluded the establishment of a reference annotation to pinpoint the existence and location of 'faults' in the original data. This makes it difficult to evaluate the algorithm's true performance. Another methodological limitation is that the database used for testing was limited to eight individuals, and the performance of the algorithm on larger populations is unknown. The technological limitations are imposed by recording and storage capabilities of man-wearable systems as well as by transmission capabilities of a local-area radio network. These systems may not be able to store or transmit the amount of information contained in ECG and respiratory waveforms effectively. The modest performance of the algorithm, in comparison with other reported results, can be attributed to the fact that the majority of these data-qualification algorithms use additional information contained in the waveforms.

In summary, we describe an algorithm to assign confidence values to HR and RR data. The algorithm is based on a fuzzy logic engine, which allows the evaluation of input features by using membership functions that are based on expert knowledge or that are extracted from physiological limits or relationships. Our method provides a feasible approach to identify usable data in noisy field-collected data streams, where it is likely that redundant measures of the vital signs will be absent. The algorithm incorporates a framework that can be easily modified to integrate new sensors as they become available, while the input feature membership functions can be adjusted to accommodate more refined estimates of the physiological relationships as they become known, or to tailor the performance of the algorithm to specific subject populations.

## Disclaimer

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the US Army or the US Department of Defense. Citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations. This manuscript has been approved for public release with unlimited distribution.

## Acknowledgments

The work was partially supported by the Combat Casualty Care and the Military Operational Medicine Directorates of the US Army Medical Research and Materiel Command, Fort Detrick, Maryland.

## References

Bailey J J, Berson A S, Garson A, Horan L G, McFarlane P W, Mortara D W and Zywietz C 1990 Recommendations for standardization and specifications in automated electrocardiography: bandwidth and digital signal processing. A report for health professionals by an ad hoc writing group of the Committee on Electrocardiography and Cardiac Electrophysiology of the Council on Clinical Cardiology, American Heart Association *Circulation* **81** 730–9

Bates J H T and Young M P 2003 Applying fuzzy logic to medical decision making in the intensive care unit *Am. J. Resp. Crit. Care Med.* **167** 948–52

Beidleman B A, Tharion W J, Buller M J, Hoyt R W and Freund B J 2004 Reliability and validity of devices of a life sign detection system *U.S. Army Research Institute of Environmental Medicine Technical Report T-05-01* (Natick, MA: US Army Research Institute of Environmental Medicine)

Cao C, Kohane I S and McIntosh N 1999 Artifact detection in cardiovascular time series monitoring data from preterm infants *Proc. AMIA Symp.* pp 207–11

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

Edworthy J and Hellier E 2006 Alarms and human behaviour: implications for medical alarms *Br. J. Anaesth.* **97** 12–7

Hoyt R W, Reifman J, Coster T S and Buller M J 2002 Combat medical informatics: present and future *Proc. AMIA Symp.* pp 335–9

Jiang J A, Chao C F, Chiu M J, Lee R G, Tseng C L and Lin R 2007 An automatic analysis method for detecting and eliminating ECG artifacts in EEG *Comput. Biol. Med.* **37** 1660–71

Macfarlane D J 2001 Automated metabolic gas analysis systems. a review *Sports Med.* **31** 841–61

Obuchowski N A 2003 Receiver operating characteristic curves and their use in radiology *Radiology* **229** 3–8

Park H J, Jeong D U and Park K S 2002 Automated detection and elimination of periodic ECG artifacts in EEG using the energy interval histogram method *IEEE Trans. Biomed. Eng.* **49** 1526–33

Shanahan J G 2000 *Soft Computing for Knowledge Discovery: Introducing Cartesian Granule Features* (Norwell, MA: Kluwer)

Sugeno M 1985 *Industrial Applications of Fuzzy Control* (New York: Elsevier)

Unnithan V B, Wilson J, Buchanan B, Timmons J A and Patton J Y 1994 Validation of the SensorMedics (S2900Z) metabolic cart for pediatric exercise testing *Can. J. Appl. Physiol.* **19** 472–9

Wahba G 1990 *Spline Models for Observational Data* (Philadelphia, PA: Society for Industrial and Applied Mathematics)

Witt J D, Fisher J R K O, Guenette J A, Cheong K A, Wilson B J and Sheel A W 2006 Measurement of exercise ventilation by a portable respiratory inductive plethysmograph *Resp. Physiol. Neurobiol.* **154** 389–95

Wolf M, Keel M, von Siebenthal K, Bucher H U, Geering K, Lehareinger Y and Niederer P 1996 Improved monitoring of preterm infants by fuzzy logic *Technol. Health Care* **4** 193–201

Xu X and Schuckers S 2001 Automatic detection of artifacts in heart period data *J. Electrocardiol.* (Suppl) **34** 205–10

Yu C, Liu Z, McKenna T, Reisner A T and Reifman J 2006 A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms *J. Am. Med. Inform. Assoc.* **13** 309–20