

LAM: A landscape matching algorithm for respiratory data alignment

Liangyou Chen, Thomas McKenna, Andrei Gribok, Jaques Reifman*
Bioinformatics Cell, U.S. Army Medical Research and Materiel Command
504 Scott Street, Ft. Detrick, MD 21702-5012

ABSTRACT

Respiratory waveforms and their derived respiratory rate time-series data can become misaligned from each other when they are collected by vital signs monitors under sub-optimal field conditions. The monitor-provided waveforms and rates can be re-aligned by independently calculating respiratory rates from the waveforms and then aligning them with the monitor-provided rates. However, substantially different rates may be generated from the same waveform due to the presence of ambiguous breaths at noisy positions in the waveform. This paper reports a landscape matching (LAM) algorithm to align respiratory rate time-series data with the waveform that they are derived from by using rates that are calculated by different means. The algorithm exploits the intermittent matches between two respiratory rate time series to generate a matching score for an alignment. The best alignment exhibits the highest matching score. The alignment performance of the LAM algorithm is compared to that of a correlation matching (CM) algorithm using field-collected respiratory data. Alignment performance is evaluated by: (1) comparing the ability of the two algorithms to return a shifted waveform to its original, known position; and (2) comparing the percent of points that match between the monitor-provided and calculated respiratory rate time-series data after re-alignment. The LAM alignment algorithm outperforms the CM algorithm in both comparisons at a statistically significant level ($p < 0.05$). Out of 67 samples with shifted time-series data, on average, the LAM aligns respiratory rates within 44 seconds of the original position, which is significantly better than the CM-calculated alignment (136 seconds). Out of 465 samples, the LAM performs better, worse, and equal to the CM algorithm in percentage of points matching in 73%, 11%, and 16% of the cases, respectively. This robust alignment algorithm supports the use of reliable *post-hoc* monitor-provided respiratory rates for data mining purposes.

Keywords: Respiratory rate, time-series alignment, correlation matching, vital signs, data mining

1. INTRODUCTION

The respiratory rate is a fundamental vital sign. It is associated with the modulation of cardiac vagal outflow and tone via respiratory sinus arrhythmia,^{1,2} and has been used to evaluate respiratory mechanics during mechanical ventilation,^{3,4} to index cardiac parasympathetic control,⁵ and to estimate trauma severity.⁶ Moreover, patterns of respiration, such as alternating respiration⁷ and Cheyne-Stokes respiration,⁸ have diagnostic value for various sleep-disorder and cardiovascular diseases.^{9,10} The simultaneous study of respiratory rate, heart rate, arterial pressure, and other physiological variables may help discover critical dynamics, trends, and events that affect patient outcomes.¹¹⁻¹⁴ Such multivariate study, however, requires the time-synchronized acquisition of physiology variables. Continuous monitoring and acquisition of respiratory rates, in synchronization with other physiological variables, is supported by various vital signs monitors, and tools have been developed to record respiratory waveforms and their derived rate data and export them into external storage media.^{12,15} The availability of these tools and techniques enables large-scale data warehousing and the *post-hoc* study of field-collected physiological time-series data,¹⁶ which is essential for the development of automated diagnostic and/or prognostic algorithms to assist in emergency care and combat casualty monitoring systems.

Respiratory waveforms collected in the field are noisy, and result in erratic derived respiratory rates. Our *post-hoc* study of physiology data records from 735 trauma patients, collected during transport from the field to hospital, shows that data quality is a major issue that must be addressed to support the development of reliable diagnostic or prognostic algorithms for trauma patient care.¹⁷ Misaligned, missing and mislabeled data are common, possibly because of

* jaques.reifman@us.army.mil; phone 1-301-619-7915; fax 1-301-619-1983

interruptions between the recording of the respiratory waveforms and the rates under the relatively unstructured environment that pertains during transport of patients, or errors in the data archiving process. Assuring the proper alignment of physiology time-series and quantifying their quality is necessary for data-mining purposes. This paper presents an algorithm to align low-quality respiratory waveforms, and their associated, but possibly time-misaligned rates, with high accuracy. The algorithm independently recalculates the respiratory rate from the respiratory waveform, and then compares the calculated respiratory rate (RR_C) to the monitor-provided reference respiratory rate (RR_R) to determine if temporally repositioning one series with a certain shift will result in a better match between the two respiratory rate series, and, hence, estimating the necessary shift to correct the alignment between the monitor-provided waveform and RR_R .

The alignment problem is challenging because respiratory rates, derived from the same waveform segment, can vary greatly depending on how breaths are counted. For example, a 15-second respiratory waveform from a typical patient breathing at 20-breaths/minute contains 5 breaths, which are identified by counting the number of breath cycles in the respiratory waveform. However, a noisy respiratory waveform may include a mixture of deep and shallow breaths, in which case deciding to include or exclude potential breaths for breath-counting purposes becomes highly subjective, even for human experts.^{12,18} Therefore, differently implemented breath-identification algorithms may generate substantially different respiratory rates. For instance, missing one breath cycle from the above 15-second respiratory waveform will yield a 16-breaths/minute respiratory rate or a 20% difference from the 'actual' rate. Respiratory waveforms collected in the field are especially noisy which, in combination with other effects, such as artifacts generated by subject movement and by poor placement of sensor electrodes on the body,¹⁹ lead to additional ambiguities in breath identification and to a greater disagreement in respiratory rates as calculated by different algorithms. An alignment algorithm that can tolerate such noisy signals and large outliers is, therefore, desirable.

Alignment of the respiratory data can be accomplished by employing a matching score, which measures the cumulative distance between two respiratory rate time-series, with a certain shift applied to one series. The shift that maximizes the matching score between the two aligned series is deemed the best alignment. We found that our alignment algorithm, termed LAM (Landscape Matching), which employs a landscape matching protocol, is more sensitive and accurate than an algorithm that uses correlation matching. The LAM algorithm capitalizes on alignments that occur among pair-wise points of two time-series by employing a parameter that strengthens the simultaneous change of values in neighboring matching positions. This method, therefore, tolerates large outliers while emphasizing intermittent matches between the two respiratory rate time series. Such functionality is lacking in other functions which measure distance between time series, such as the average Euclidean, Manhattan, and Minkowski distances between all pairs of points.²⁰ Adoption of the LAM algorithm significantly improves the accuracy of the alignment of our respiratory data.

We describe, in Section 2, the respiratory waveform and respiratory rate data on which our algorithm is developed and tested. The correlation matching and LAM algorithms are detailed in Section 3. Their performance is compared in Section 4, followed by conclusions in Section 5.

2. DATA

Complete vital signs records for 735 trauma injured patients are warehoused in our system for the management and analysis of time-series physiological data, which is termed the Physiology Analysis System (PAS).¹⁶ These data were collected from patients during transport by Life Flight helicopter service from the site of injury to the Memorial Hermann Hospital, an urban regional trauma center at Houston, Texas.²¹ Vital signs data were collected during flight with a Propaq Encore® 206EL (Software Version 2.0X) vital signs monitor;¹⁵ the data were then exported onto a personal digital assistant flashcard, uploaded to a local database²² and then exported into the PAS.

The Propaq monitor collects respiratory waveforms by impedance pneumography from the same leads used to collect electrocardiogram (ECG) waveform data. The monitor determines which signals are artifacts and which signals result from actual respiratory efforts. The ECG waveform is collected at 182 Hz and the respiratory waveform is collected at 22.7 Hz. The ECG heart rate and the respiratory rate are derived from the ECG waveform and respiratory waveform, respectively, and both are provided at 1.0 Hz intervals.¹⁵ Each monitor-provided respiratory rate and heart rate datum is recorded along with a time stamp corresponding to the actual time the datum is read, while each respiratory waveform

and ECG waveform datum is recorded with a sequential number starting from 1. Thus, if the derived respiratory rates are not aligned with the respiratory waveform, we assume that the same misalignment holds between the derived heart rates and the ECG waveform. With this working assumption, we verify the accuracy of our alignment algorithm for the respiratory rate by comparing it with the alignment of good-quality ECG waveform and heart rate data. This is due to the fact that ECG heart beats can be identified without ambiguity from good-quality ECG waveforms, and, therefore, alignment of recalculated heart rates from good-quality ECG waveforms and the device-provided heart rates can be much more reliably determined. It is worth mentioning that the main purpose of our respiratory data alignment algorithm is to robustly realign respiratory data when the corresponding ECG waveform is unavailable or is very noisy and the corresponding ECG heart rate is hard to compute. Patients with good-quality ECG waveforms are selected by application of an in-house developed heart rate quality index algorithm.¹⁷ Therefore, we use the heart rate alignment shifts from good-quality ECG waveforms as a ‘gold standard,’ and as the ‘known’ respiratory shifts to evaluate the respiratory data alignment algorithm.

From the total population of 735 patients, a test population of 67 patients was selected, each with ≥ 2 minutes of continuous ECG waveform, and we required that $\geq 70\%$ of each ECG waveform series was of good quality. The ECG waveforms were previously aligned to the ECG heart rates,¹⁷ and the expected equivalent alignment shift for each respiratory rate was calculated. For each of the 67 patients, a segment of respiratory waveform is extracted from the original waveform at a random starting position and the length of the extracted waveform segment is selected by multiplying the total waveform length by a ratio r , which is chosen in a range between 0.5 and 1.0. The recalculated RR_C , the reference RR_R , and the expected alignment position, which is determined by the referred ECG alignment shift plus the random starting position, for each extracted waveform segment are composed as a test sample. The 67 test samples, each from a different patient, constitute one test dataset. The process was repeated five times by fixing r , and repetitively sampling from different random starting positions of each patient's respiratory waveform. The five datasets associated with a fixed r are termed an r -collection. We then generated 6 r -collections using r values of 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0, for a total of 2010 ($5 \times 6 \times 67$) test samples.

3. METHODOLOGY

This section describes two algorithms to align respiratory rates. The first, the correlation matching (CM) algorithm, utilizes the correlation coefficient of two time series at each aligned position to calculate a matching score. The second, the LAM algorithm, utilizes a landscape vector, λ , to calculate the matching score. Before detailing the algorithms we first introduce some concepts.

3.1. Definition of concepts

Let a *time series* of length T be represented by vector \mathbf{x} , whose value at position t , $t = 1, 2, \dots, T$, is denoted as $x(t)$. A shifted time series \mathbf{x}_{+s} of \mathbf{x} , with a *shift* s , is a time series whose value at position t is $x_{+s}(t) = x(t - s)$, for all valid t . The overlapping segment of two time series, \mathbf{x}_1 and \mathbf{x}_2 , of length T form sub-series of length T^* , with $T^* \leq T$, denoted as \mathbf{x}_1^* and \mathbf{x}_2^* , respectively. These two time series, derived from the same waveform, are *misaligned* if there is a shift $s_0 \neq 0$, such that \mathbf{x}_1 and $\mathbf{x}_{2,+s_0}$ have an overlapping segment whose values are derived from the same positions in the waveform. Given two misaligned time series, the alignment problem is to find their unknown original shift s_0 . Throughout this paper, we assume that the shift will be applied to the second time series in the order of the input arguments, and that the first time series will maintain its original position. We further assume that the shift is positive, in that the second time series is shifted by $+s$ positions from $t - s$ to t . However, our alignment method is general, and it is equally suitable for data requiring negative shifts.

We estimate the optimal alignment shift \hat{s} between two time series, \mathbf{x}_1 and \mathbf{x}_2 , as the shift s that maximizes a matching function $M_s(\mathbf{x}_1, \mathbf{x}_{2,+s})$, that is, $\hat{s} = \arg \max_s [M_s(\mathbf{x}_1, \mathbf{x}_{2,+s})]$, where $M_s(\mathbf{x}_1, \mathbf{x}_{2,+s})$ is a correlation-based function M_s^C or a LAM function $M_s^{L,\theta}$ discussed below.

The performance of an estimated alignment shift \hat{s} is evaluated by calculating the difference $D(\hat{s}, s_0)$ from the known shift s_0 , where $D(\hat{s}, s_0) = |\hat{s} - s_0|$. A smaller difference represents a better alignment and the ideal alignment yields

$D(\hat{s}, s_0) = 0$, representing the condition where the estimated alignment shift matches the actual shift exactly. The difference D requires the known actual alignment shift s_0 to be known.

Another method to evaluate an alignment is to count the number of matched positions of the aligned time series. Let \mathbf{x}_1 and \mathbf{x}_2 be two time series of length T_1 and T_2 , respectively, and θ be a matching threshold. The *matching positions* of \mathbf{x}_1 and \mathbf{x}_2 under threshold θ are defined as all positions t that satisfy

$$|x_1(t) - x_2(t)| \leq \theta. \quad (1)$$

The matching positions sorted in the order of time constitute a time series denoted as $\tau_{\mathbf{x}_1, \mathbf{x}_2}^\theta$ of length T_θ , where $\tau_{\mathbf{x}_1, \mathbf{x}_2}^\theta(k)$, $k=1, 2, \dots, T_\theta$, denotes the time of the k 'th matching position. The *percentage match* (PM) of an estimated alignment shift \hat{s} , for the alignment of \mathbf{x}_1 and \mathbf{x}_2 under threshold θ , is defined as:

$$\text{PM}_{\mathbf{x}_1, \mathbf{x}_2}^\theta(\hat{s}) = \frac{T_\theta}{T_2} \times 100\%. \quad (2)$$

PM ranges from 0% to 100%, and a larger PM value stands for a potentially better alignment because more positions are matched under the threshold θ .

3.2. The correlation matching (CM) algorithm

The correlation matching algorithm employs a correlation function to quantify the alignment between \mathbf{x}_1 and \mathbf{x}_2 . The correlation matching function M_s^C is defined as the product of the correlation coefficient²³ of the overlapping segment of the two time series and the ratio formed by dividing the length of the overlapping time series by the length of the second time series. Given a time series \mathbf{x}_1 of length T_1 and a second time series $\mathbf{x}_{2,+s}$ of length T_2 shifted by s positions and overlapping over T^* positions, the M_s^C is calculated by

$$M_s^C(\mathbf{x}_1, \mathbf{x}_{2,+s}) = \frac{T^*}{T_2} \cdot \frac{\text{cov}(\mathbf{x}_1^*, \mathbf{x}_{2,+s}^*)}{\sqrt{\text{cov}(\mathbf{x}_1^*, \mathbf{x}_1^*) \text{cov}(\mathbf{x}_{2,+s}^*, \mathbf{x}_{2,+s}^*)}}, \quad (3)$$

where

$$\text{cov}(\mathbf{y}_1, \mathbf{y}_2) = \frac{1}{T-1} \sum_{i=1}^T \{[y_1(i) - \bar{y}_1] \cdot [y_2(i) - \bar{y}_2]\} \quad (4)$$

denotes the covariance of two time series \mathbf{y}_1 and \mathbf{y}_2 of an equal length T , and \bar{y}_1 and \bar{y}_2 denote the mean values of \mathbf{y}_1 and \mathbf{y}_2 , respectively. The ratio T^*/T_2 in Equation (3) penalizes short overlaps between the two time series. Hence, a better correlation and a larger overlap between \mathbf{x}_1 and $\mathbf{x}_{2,+s}$ result in a larger M_s^C . M_s^C ranges from -1 to 1 , with a higher value standing for a better match between the two time series.

The CM algorithm evaluates M_s^C for all potential alignment shifts between RR_R , of length T_R , and RR_C , of length T_C , and estimates the best alignment shift \hat{s} associated with the largest M_s^C . For example, for respiratory rate time series $\text{RR}_R = (20, 26, 27, 18, 22)$ and $\text{RR}_C = (20, 26, 18, 19, 22)$ shown in Figure 1(a), the matching scores M_s^C are 0.31, 0.62, -0.40 and 0.40 for alignment shifts s of 0, 1, 2 and 3, respectively. The largest matching score of 0.62 is obtained for $s = 1$. As a result, a $\hat{s} = 1$ means that RR_C should be shifted to the right by 1 second to match RR_R (as illustrated in Figure 1(b)). Given a matching threshold $\theta = 2$, two positions, 3 and 4, are matched with the alignment shift $\hat{s} = 1$, with a corresponding percentage match of 40%.

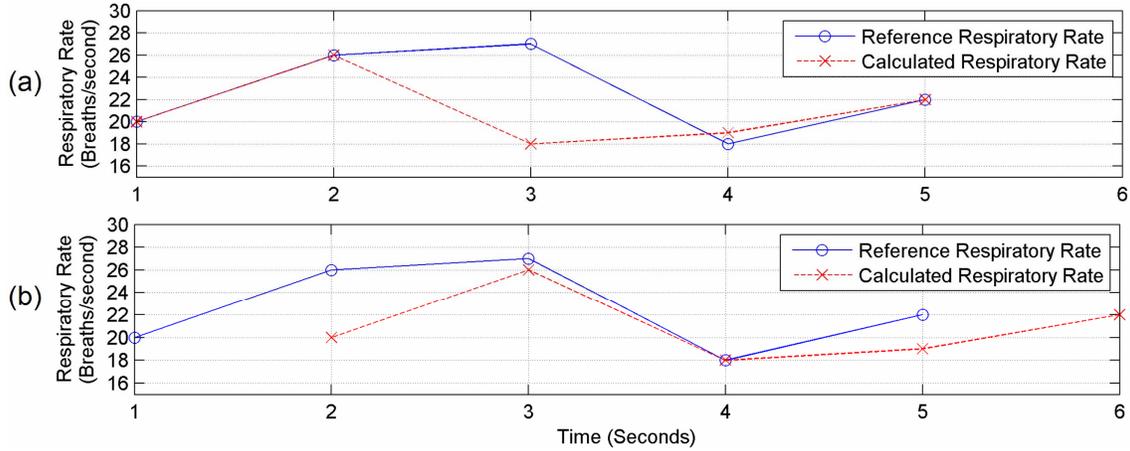


Figure 1. Alignment of two respiratory rate time series. (a) Unaligned reference and calculated respiratory rates. (b) The alignment of respiratory rates from (panel a) after the calculated respiratory rate is shifted to the right by 1 second.

3.3. The landscape matching (LAM) algorithm

The landscape matching algorithm employs a landscape function to quantify the alignment between \mathbf{x}_1 and \mathbf{x}_2 . Let \mathbf{x}_1 and \mathbf{x}_2 be two time series of lengths T_1 and T_2 , respectively, and θ be a matching threshold, defining the *landscape vector* $\lambda_{\mathbf{x}_1, \mathbf{x}_2}^\theta$, whose value at position t is defined by

$$\lambda_{\mathbf{x}_1, \mathbf{x}_2}^\theta(t) = \begin{cases} |x_1(t) - x_1(t')|, & \text{where } \begin{cases} t = \tau_{\mathbf{x}_1, \mathbf{x}_2}^\theta(k), \\ t' = \tau_{\mathbf{x}_1, \mathbf{x}_2}^\theta(k-1), \end{cases} \text{ for } 1 < k \leq T_\theta \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where, as above, $\tau_{\mathbf{x}_1, \mathbf{x}_2}^\theta(k)$, $k=1, 2, \dots, T_\theta$ denotes the time of the k 'th matching position between \mathbf{x}_1 and \mathbf{x}_2 defined by Equation (1). In other words, if t is the k 'th matching position of \mathbf{x}_1 and \mathbf{x}_2 under threshold θ , with $t = \tau_{\mathbf{x}_1, \mathbf{x}_2}^\theta(k)$ for $1 < k \leq T_\theta$, then $\lambda(t)$ is assigned the absolute difference of \mathbf{x}_1 's values at the k 'th matching position, $x_1(t)$, and the $(k-1)$ 'th matching position, $x_1(t')$, where $t' = \tau_{\mathbf{x}_1, \mathbf{x}_2}^\theta(k-1)$; otherwise $\lambda(t)$ is set to 0. Hence, λ can be intuitively thought as the absolute change in amplitude (or landscape) of two successive matching (not necessary sequential) positions in \mathbf{x}_1 . For example, consider the respiratory rate time series $RR_R = (20, 26, 27, 18, 22)$ and $RR_C = (20, 26, 18, 19, 22)$ from the previous section with $\theta = 2$. Aligning these two time series with no shifts yields four matching positions 1, 2, 4 and 5, comprising $\tau_{RR_R, RR_C}^{\theta=2} = (1, 2, 4, 5)$ associated with RR_R values of 20, 26, 18 and 22, respectively. The absolute changes of RR_R at these four matching positions are 0, 6, 8 and 4, respectively, where the first matching position is set to zero according to Equation (5). The landscape vector is, therefore, $\lambda_{RR_R, RR_C}^{\theta=2} = (0, 6, 0, 8, 4)$.

Given a matching threshold θ , the landscape matching function $M_s^{L, \theta}$ for the alignment of time series \mathbf{x}_1 of length T_1 and \mathbf{x}_2 of length T_2 with an alignment shift s is defined as

$$M_s^{L, \theta}(\mathbf{x}_1, \mathbf{x}_{2, +s}) = \frac{1}{T_1} \sum_{t=1}^{T^*} \frac{1 + \lambda_{\mathbf{x}_1^*, \mathbf{x}_{2, +s}^*}^\theta(t)}{1 + |x_1^*(t) - x_{2, +s}^*(t)|}, \quad (6)$$

where the label L stands for the landscape matching algorithm. $M_s^{L, \theta}$ is a function that measures the difference between \mathbf{x}_1^* and $\mathbf{x}_{2, +s}^*$ and is adjusted by the elements of the landscape vector λ . When \mathbf{x}_1^* and $\mathbf{x}_{2, +s}^*$ match each other perfectly, $|x_1^*(t) - x_{2, +s}^*(t)| = 0$ for all t , and the value of $M_s^{L, \theta}$ reflects the summation of the absolute changes in the amplitude of successive \mathbf{x}_1^* and $\mathbf{x}_{2, +s}^*$ values, i.e., the elements of the landscape vector λ , over the length of the

overlapping time series and, therefore, $M_s^{L,\theta}$ can be arbitrarily large. Conversely, when x_1^* and $x_{2,+s}^*$ are far apart and $|x_1^*(t) - x_{2,+s}^*(t)| > \theta$ for all t , the elements of λ will be zero and $M_s^{L,\theta}$ can be arbitrarily small. Thus, $M_s^{L,\theta} > 0$ and has no upper bound. A larger $M_s^{L,\theta}$ value reflects a closer match between two time series. The $M_s^{L,\theta}$ function also depends on the selection of the parameter θ , which is the threshold at which we consider two points to match each other. It is useful to choose a θ that has a practical meaning. For example, expert opinion suggests that a difference of 2 breaths/min for the respiratory rate is acceptable for a match, and we use $\theta = 2$ breaths/min as a matching threshold to align respiratory data in this paper.

The LAM algorithm evaluates $M_s^{L,\theta}$ for all potential alignment shifts between RR_R of length T_R and RR_C of length T_C using a fixed θ , and estimates the best alignment shift \hat{s} associated with the largest $M_s^{L,\theta}$. For example, for the previous respiratory rate time series $RR_R = (20, 26, 27, 18, 22)$ and $RR_C = (20, 26, 18, 19, 22)$ and $\theta = 2$, the matching scores $M_s^{L,2}$ are 3.52, 2.18, 0.09, 0.11 and 0.07 for alignment shifts s equal to 0, 1, 2, 3 and 4, respectively. The largest $M_s^{L,2}$ of 3.52 is obtained for $s = 0$. As a result, $\hat{s} = 0$, means that the best match between RR_C and RR_R is the one requiring no shift, as illustrated in Figure 1(a). With the matching threshold $\theta = 2$, four positions, 1, 2, 4 and 5, are matched, yielding a percentage match PM of 80%. In this example, the LAM algorithm aligns these two respiratory time series with a higher PM than the CM algorithm, which attained a PM of 40%.

4. RESULTS

We compare the performance of the LAM and CM algorithms by: (1) comparing the alignment results of a typical patient; (2) comparing the ability of the two algorithms to return shifted waveforms to their original positions for the datasets of the test population; and (3) comparing the algorithms' respective percentage match when aligning the respiratory rates for the entire population of patients.

4.1. Comparison of LAM and CM alignments from a typical patient

We compare the optimal alignment provided by the LAM and the CM algorithm by using the respiratory rate taken from a typical patient (#757). Figure 2 plots the matching scores $M_s^{L,2}$ and M_s^C for the LAM and CM algorithms, respectively, after alignment of the respiratory rate at all possible shift positions. The matching score $M_s^{L,2}$ for the LAM algorithm has a maximum value of 0.44 at a shift of 431 seconds, while the matching score M_s^C for the CM algorithm has a maximum value of 0.33 at a shift of 40 seconds. The percentage matches $PM^{(\theta=2)}$ after the shifts are 70% and 44%, respectively, for the LAM and the CM algorithm, meaning that the alignment obtained by the LAM algorithm results in a 26% increase in the number of matching positions when compared with the CM alignment. Figure 3 shows the alignment of the RR_R and RR_C for patient #757 after the shift by each algorithm. It is clear that RR_R and RR_C match better after a 431-second shift of RR_C by the LAM algorithm (lower panel) than the 40-second shift suggested by the CM algorithm in the top panel.

In order to examine how the magnitude of θ influences the function of the LAM algorithm, we repetitively calculated matching scores $M_s^{L,\theta}$ using different values of θ for the same patient (Figure 4). The results show that although the amplitudes of the results are slightly different, their shapes remain very similar, all yielding the same optimal shift for this patient. This example shows that the LAM algorithm is relatively insensitive to the choice of different θ . It is worth mentioning, however, that LAM may generate different optimal shifts using different θ for other patients. The selection of different θ also affects the percentage match PM function, which is used to evaluate the alignments by the LAM and the CM algorithms. We employed different θ matching thresholds of 1, 3, 4 and 5 and calculated the respective PM for the previous alignments of patient #757 by both algorithms; the corresponding PMs are 46/27, 81/58, 86/68 and 89/75, respectively, for the LAM/CM alignment, indicating that the LAM alignment yields a consistently better PM than that obtained by CM, regardless of the choice of θ .

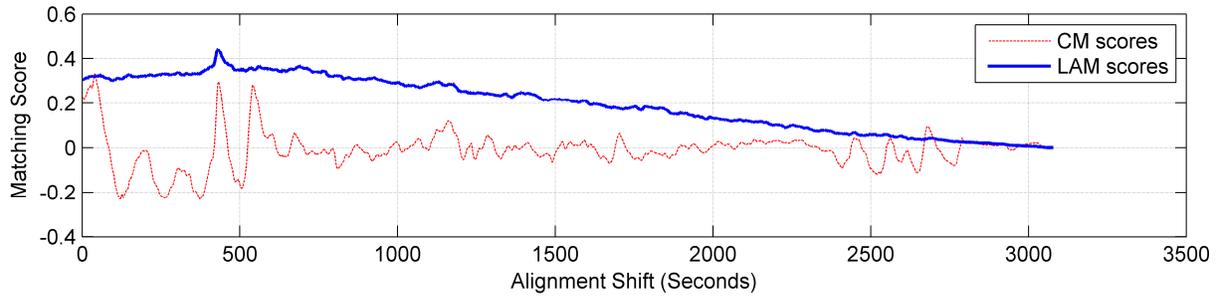


Figure 2. Tracings of matching scores at all possible shift positions after alignment of respiratory rates for patient 757 by the LAM and the CM algorithms. The LAM algorithm yields a peak score of 0.44 at a shift of 431 seconds, while the CM algorithm yields a peak score of 0.33 at a shift of 40 seconds.

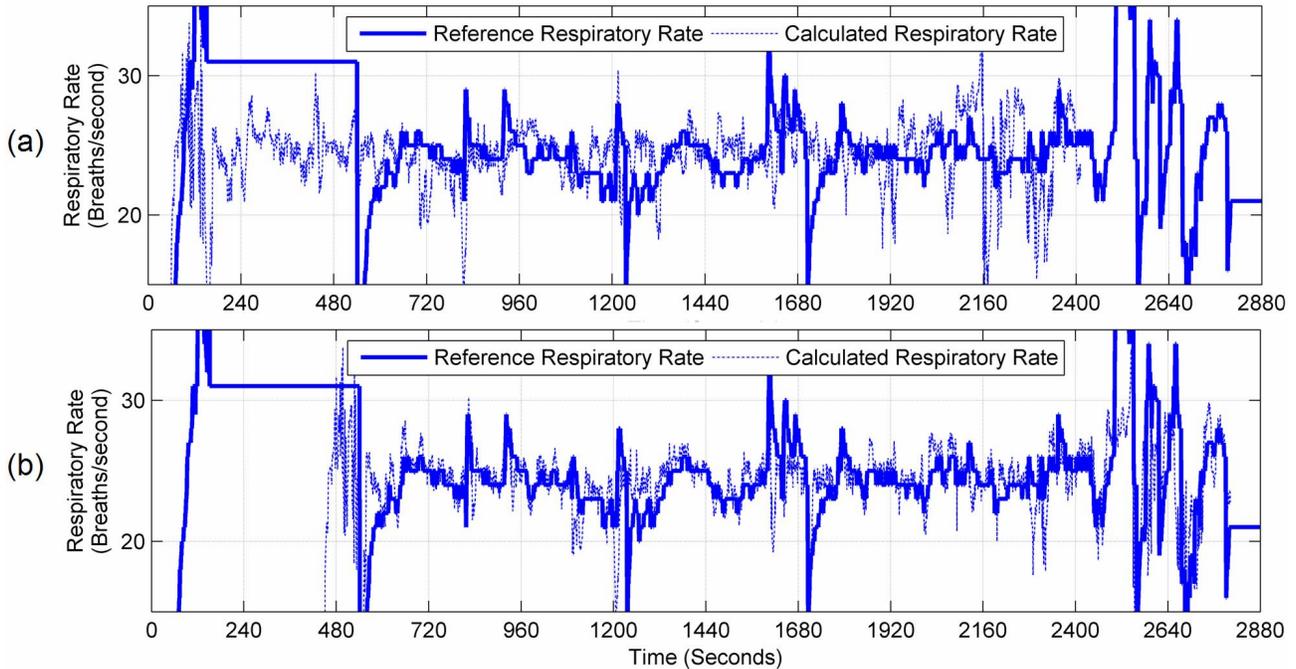


Figure 3. The alignment of reference and calculated respiratory rates by the CM (panel a) and LAM (panel b) algorithms. (a) Plot of reference and calculated respiratory rates after a 40-second shift of the calculated respiratory rate, as determined by the CM algorithm. Forty-four percent of the calculated respiratory rates match the reference respiratory rates within 2 breaths/min after the shift. (b) Plot of reference and calculated respiratory rates after a 431-second shift of the calculated respiratory rate, as determined by the LAM algorithm. Seventy percent of the calculated respiratory rates match the reference respiratory rates within 2 breaths/min after the shift.

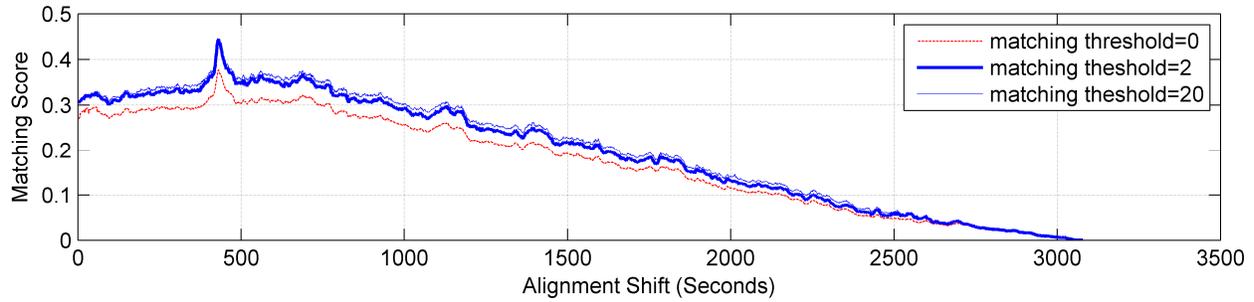


Figure 4. Tracings of matching scores $M_s^{L,\theta}$ for the alignment of respiratory rates for patient 757 by the LAM algorithm using different matching thresholds θ . The shapes of the curves are very similar, though their amplitudes are different, and they all find the same shift for this patient. This result shows that, for this patient, the LAM algorithm is relatively insensitive to the choice of θ .

4.2. Comparison of LAM and CM alignments with a known shift

We compared the performance of the LAM and CM algorithms by computing the difference $D(\hat{s}, s_0) = |\hat{s} - s_0|$ provided by each algorithm in the alignment between RR_R and RR_C for the test population of 67 patients, where the alignment shift s_0 was known for each time-series record. The best alignment algorithm is the one yielding the smallest mean value of D over the test dataset.

For each of the five test datasets in an r -collection, we aligned all samples with the LAM and the CM algorithms and calculated their D values. The 67 D values obtained by the LAM algorithm for a test dataset are compared to the respective 67 D values obtained by the CM algorithm, and the differences are tested for statistical significance using the Student's T-test. The mean and standard deviation (Std) of the 67 D values and the p -values of the Student's T-test are averaged over the five datasets in an r -collection and the averaged mean, Std, and p -value corresponding to each r , ranging from 0.5 to 1.0, are shown in Table 1. The results show that for relatively short RR_C segments (average <1140 seconds) obtained for $r < 0.8$, the LAM algorithm generates alignments that yield smaller mean values of D , although the difference are not statistically significant for a $p < 0.05$. However, for longer RR_C segments, $r \geq 0.8$, the LAM algorithm generates alignments significantly ($p < 0.05$) closer to the known shift than the CM algorithm.

With $r = 1$, which means that the total length of RR_C is used for the alignment with RR_R , the LAM algorithm yields a mean difference D of 44 seconds, which is acceptable considering the presence of some large D -value outliers inflating the mean results; the comparable mean D for the CM algorithm is 136 seconds. When $r = 1$, the LAM and the CM algorithm align 45 (67%) and 37 (55%) of the 67 test patients with D values <15 seconds, respectively, and align 54 (80%) and 41 (61%) patients with D values <30 seconds, respectively. Therefore, more accurate alignments can be obtained by using the LAM algorithm than by using the CM algorithm.

Table 1. Comparison of alignment performance by the LAM and CM algorithms using the test datasets with known shifts from 67 patients.

r	Average length of the sampled RR_C (seconds)	Mean \pm Std of the difference $D = \hat{s} - s_0 $ (seconds, averaged over 5 test-datasets in the r -collection)		p -value of Student's T-test (averaged over 5 test-datasets in the r -collection)
		LAM	CM	
0.5	712	184 ± 260	195 ± 295	0.60
0.6	855	157 ± 230	194 ± 298	0.42
0.7	997	112 ± 187	162 ± 253	0.20
0.8	1140	81 ± 140	174 ± 285	0.03
0.9	1282	54 ± 113	156 ± 260	0.00
1.0	1425	44 ± 108	136 ± 264	0.01

4.3. Comparison of LAM and CM alignments by percentage match

Respiratory data from all 735 patients were aligned by the LAM and CM algorithms. The average lengths \pm the standard deviation for RR_R and RR_C are 1543 ± 764 and 1367 ± 757 seconds, respectively. The percentage match PM after respiratory rate alignments was calculated, with the best alignment algorithm providing a larger PM. The average PM for the LAM algorithm is 0.18 ± 0.21 , and it is 0.13 ± 0.17 for the CM algorithm. This difference being statistically significant at a $p<0.05$ level with a Student's T-test. There are 484 (66%), 173 (23%) and 78 (11%) patients that have PM values after the LAM alignment that are, respectively, greater than, equal to, and less than the PM values after the CM alignment. A subgroup of 465 patients had M_S^C and $M_S^{L,2}$ scores that were both greater than 0.1 and were considered to have acceptable respiratory rate alignment by each algorithm. For this subgroup, we computed the PM values for each alignment algorithm and compared their results. The average PM, as calculated by the LAM algorithm, is 0.23 ± 0.18 , while the average PM value calculated by the CM algorithm is 0.19 ± 0.19 ($p<0.05$, Student's T-test). There are 341 (73%), 50 (11%) and 74 (16%) patients that have PM from the LAM alignment that are, respectively, greater than, equal to, and less than the PM from the CM alignment. The LAM algorithm significantly outperforms the CM algorithm in all of the PM comparisons, with a better PM for more than 66% of the patients.

5. CONCLUSION

We developed a landscape matching LAM algorithm to re-align respiratory waveform and respiratory rate time-series data that become misaligned during data recording and/or data manipulation processes. The algorithm tolerates large differences between calculated respiratory rates derived from the same respiratory waveform by different respiratory rate calculation algorithms, and accounts for the intermittent matches that occur between the differently-calculated respiratory rate time-series data. Employing multiple performance metrics, the LAM algorithm exhibited consistent and significantly better re-alignment performance than an alignment algorithm based on correlation coefficient. The LAM algorithm provides a reliable method for the realignment of noisy respiratory data.

ACKNOWLEDGEMENTS

We are grateful to the University of Texas Health Sciences Center, and to COL John Holcomb and Dr. Jose Salinas of the U.S. Army Institute of Surgical Research, Fort Sam Houston, San Antonio, Texas, for providing access to the vital signs database. We thank Gagandeep Bawa, Kamal Kumar and the Physiology Analysis System team for providing a data analysis platform for trauma patients. Some of the respiratory data alignment computations were performed on high-performance computers at the U.S. Army Research Laboratory, Major Shared Resource Center, sponsored by the Department of Defense High Performance Computing Modernization Program. This work was supported by the Combat Casualty Care and the Military Operational Medicine Directorates at the U.S. Army Medical Research and Materiel Command, Fort Detrick, Maryland.

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 U.S. Army or the U.S. Department of Defense.

REFERENCES

1. G. G. Berntson, J. T. Cacioppo and K. S. Quigley, Respiratory sinus arrhythmia: autonomic origins, physiological mechanisms, and psychophysiological implications, *Psychophysiology*, 30(2):183-96, 1993.
2. F. Yasuma and J. Hayano, Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm? *Chest*, 126(4):1385-6, 2004.
3. S. C. Yang and S. P. Yang, Effects of inspiratory flow waveforms on lung mechanisms, gas exchange and respiratory metabolism in COPD patients during mechanical ventilation, *Chest*, 122:2096-104, 2002.
4. U. Lucangelo, F. Bernabe and L. Blanch, Respiratory mechanics derived from signals in the ventilation circuit, *Respiratory Care*, 50(1):55-67, 2005.

5. P. Grossman, F. H. Wilhelm and M. Spoerle, Respiratory sinus arrhythmia, cardiac vagal control, and daily activity, *Am. J. Physiol. Heart Circ. Physiol.*, 287(2):H728-34, 2004.
6. S. P. Frankema, E. W. Steyerberg, M. J. Edwards and A. B. van Vugt, Comparison of current injury scales for survival chance estimation: an evaluation comparing the predictive performance of the ISS, NISS, and AP scores in a Dutch local trauma registration, *J. Trauma*, 58(3):596-604, 2005.
7. R. J. Thomas, M. G. Terzano, L. Parrino and J. W. Weiss, Obstructive sleep-disordered breathing with a dominant cyclic alternating pattern--a recognizable polysomnographic variant with practical clinical implications, *Sleep*, 27(2):229-34, 2004.
8. R. Lorenz and A. Ito, The definition of "Cheyne-Stokes rhythms", *Acta Neurochir (Wien)*, 43(1-2):61-76, 1978.
9. Iu. V. Isakov, I. V. Krylova and Iu. S. Ioffe, Potentials of hyperbaric oxygenation in correcting central respiratory disorders in ischemic stroke, *Zh Vopr Neurokhir Im N N Burdenko*, (3):22-8, 1979.
10. J. P. Bounhoure, M. Galinier, A. Didier and P. Leophonte, Sleep apnea syndromes and cardiovascular disease, *Bull. Acad. Natl. Med.*, 189(3):445-59, 2005
11. P. Burdett-Smith, A patient who changed my practice: always check the respiratory rate, *Br. Med. J.*, 314:1549, 1997.
12. M. Folke, L. Cernerud, M. Ekstrom and B. Hok, Critical review of non-invasive respiratory monitoring in medical care, *Med. Biol. Eng. Comput.*, 41(4):377-83, 2003.
13. J. Haba-Rubio, G. Darbellay, F. R. Herrmann, J. G. Frey, A. Fernandes, J. M. Vesin, J. P. Thiran and J. M. Tschoop, Obstructive sleep apnea syndrome: effect of respiratory events and arousal on pulse wave amplitude measure by photoplethysmography in NREM sleep. *Sleep Breath*, 9(2):73-81, 2005.
14. U. Zwiener, C. H. Schelenz, S. Bramer and D. Hoyer, Short-term dynamics of coherence between respiratory movements, heart rate, and arterial pressure fluctuations in severe acute brain disorders, *Physiol. Res.*, 52(5):517-24, 2003.
15. Protocol® Systems Inc, *Propaq Encore® Vital Signs Monitor Reference Guide – Models 202EL, 204EL, 206EL Software Version 2.0X*, Welch Allyn, Beaverton, Oregon, 1998.
16. J. Reifman, J. Gunawardena and Z. Liu, PhysioQuest: Management and analysis of physiological data, 4th *IEEE Int. Symp. on Signal Processing and Information Technology*, pp. 521-525, Vittorio Veneto, Rome, Italy, Dec. 18-21, 2004.
17. C. Yu, Z. Liu, T. McKenna, A. T. Reisner and J. Reifman, A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms, *JAMIA*, 13(3) May/June, 2006.
18. S. M. Burns, Working with respiratory waveforms: how to use bedside graphics, *AACN Clin. Issues.*, 14(2):133-44, 2003.
19. N. D. Khambete, B. H. Brown and R. H. Smallwood, Movement artefact rejection in impedance pneumography using six strategically placed electrodes, *Physiol. Meas.*, 21(1):79-88, 2000.
20. A. R. Webb, *Statistical Pattern Recognition, Second Edition*, John Wiley & Sons, Inc., West Sussex, England, 2002. pp. 419-25.
21. V. A. Convertino and J. B. Holcomb, Advanced diagnostics for the combat medic, *US Army Medical Department Journal*, pp. 42-48, Jul.-Sep., 2003.
22. Trauma Vitals, <http://traumavitals.tamu.edu/>, *U.S. Army Institute of Surgical Research*, 2004.
23. N. R. Draper and H. Smith. *Applied regression analysis, Third Edition*, John Wiley & Sons, Inc, 1998. pp. 40-4.