



**HAL**  
open science

## **Artifact rejection and cycle detection in immature breathing: Application to the early detection of neonatal sepsis**

X Navarro, F Porée, A Beuchée, G Carrault

### ► **To cite this version:**

X Navarro, F Porée, A Beuchée, G Carrault. Artifact rejection and cycle detection in immature breathing: Application to the early detection of neonatal sepsis. *Biomedical Signal Processing and Control*, 2015, 16, pp.9 - 16. <10.1016/j.bspc.2014.10.007>. <hal-01082989>

**HAL Id: hal-01082989**

**<https://hal.science/hal-01082989v1>**

Submitted on 14 Nov 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Artifact rejection and cycle detection in immature breathing: application to the early detection of neonatal sepsis

X. Navarro<sup>1,2\*</sup>, F. Porée<sup>1,2</sup>, A. Beuchée<sup>1,2,3</sup>, G. Carrault<sup>1,2</sup>

<sup>1</sup>INSERM, U1099, Rennes, F-35000, France

<sup>2</sup>Université de Rennes 1, LTSI, Rennes, F-35000, France

<sup>3</sup>CHU Rennes, Pôle Médico-Chirurgical de Pédiatrie et de Génétique Clinique, Rennes, F-35000, France

## Abstract

This paper proposes a new framework to obtain quality respiratory variability signals from the raw breathing recorded in neonatal intensive care units (NICUs). It combines three consecutive blocks: an automatic rejection of artifacts, implemented by a logistic regression classifier, a two-step filtering process, and the identification of respiratory cycles, implemented by a peak detection algorithm. By means of a gold standard built from a preterm infants database, the performances of the first and third blocks have been evaluated. While the former obtains a 86% of specificity and sensitivity, the latter attains a respective 97%. The interest of our proposal in the clinical domain is illustrated by a promising application to detect promptly and non-invasively the presence of neonatal sepsis in the NICU.

**Keywords:** respiratory variability signal ; premature infant ; artifact removal ; automated detection ; logistic regression

## 1 Introduction

Preterm infants –born before 37 weeks of gestation– exhibit a very unstable breathing, typified by apneas or pauses in ventilation that may be accompanied by bradycardia, a decrease of the heart beat rhythm [1, 2]. This phenomenon, known as apnea of prematurity (AOP), is a consequence of the still underdeveloped brain and lungs, and is inversely related to the gestational age at birth [3, 4]. AOP may appear spontaneously, but it can also be provoked or become more severe when other pathologies –specially sepsis, i.e. a generalized infection– are present [5]. Regardless of their origin, sighs and respiratory pauses are the mechanism responsible for the variable manner in which the infants breathe during sleep. Typically, three different patterns can be identified: 1) Regular: quiet breathing with low variability in both amplitude and frequency. 2) Erratic: irregular breathing with high variability in both amplitude and frequency including several episodes of AOP. 3) Periodic: the alternation of pauses lasting a few seconds followed by several rapid and shallow breaths [6].

---

\*Address for correspondence:

LTSI, Campus de Beaulieu, Université de Rennes 1

263 Avenue du Général Leclerc - CS 74205 - 35042 Rennes Cedex, France

✉: x.navarro.s@gmail.com

☎: +33 2 23 23 62 20

☎: +33 2 23 23 69 17

The continuous monitoring of breathing and cardiac frequencies are of crucial importance to an early intervention and avoid or palliate the associated risks with recurrent apnea-bradycardia [7, 8]. A large effort has also been done to predict bradycardia [9] and to detect sepsis from the analysis of heart rate series [10, 11], but the respiratory signal has retained less attention. In the present work, breathing signals acquired in neonatal intensive care units (NICUs) are properly processed so that they can be further analyzed to add more insights about the pathological state of the premature infant. As raw signals, provided by abdominal strain gauges, are uncalibrated and cannot be used to study the air flow, they are converted to respiratory variability series (RVS). This data describes the respiratory rhythm by simply sequencing the time duration of breaths and holds interesting properties, such as long-range dependence [12].

This paper is organized as follows: Section 2 presents the database, composed by the breathing traces and clinicians' manual marks (or 'gold standard') and the framework to obtain clean signals. Section 3 describes the evaluation methodology as well as the performance of the detection methods. In Section 4, a demonstrative example to support the interest of the here-proposed framework is reported. Finally, a conclusion is drawn in the last section.

## 2 Material and Methods

### 2.1 Data selection

The breathing signals employed in this work have been selected from a larger database, collected at the University Hospital of Rennes (France), already involved in previous studies [11]. Therefore, the ensemble of the 51 preterm infants served to derive two cohorts with different purposes. The first one has been employed to examine the performances of the artifact rejection and cycle detection (Validation cohort), hence it needed to be labeled manually by an experienced clinician in order to establish the references to the automated processes. The remaining group (Testing cohort) provides data to illustrate the clinical example.

Breathing was recorded by abdominal strain gauges (Pneumotrace ©, Morro Bay, USA), piezoelectric transducers responding linearly to changes in the circumference during respiration. Signals, originally sampled at  $Fs = 400$  Hz were subsequently down-sampled to 64 Hz ( $Fr$ ) after eliminating the frequency content above  $Fr/2$  by a low-pass filter (7th order Butterworth) to avoid aliasing.

#### 2.1.1 Validation cohort

This group is constituted by five preterm infants  $31.0 \pm 1.6$  weeks post-menstrual age (PMA) and  $1.06 \pm 0.29$  kg of weight. The selection was performed visually by clinicians to ensure the inclusion of different breathing patterns. A Matlab program was purposely designed to facilitate the labeling procedure to the clinician. It consisted on marking intervals of ten seconds ( $W = 10$  s) as clean (class 0) or artifacted (class 1) in a thirty-second sliding window. The ECG signal was also displayed to help the observer to make the decision. With this program, a total of 5167 marks (14.35 hours) were obtained, the 11.8% of them classified as artifacts.

In a second instance, artifact-free marked periods were used to build the references for the automatic detection of breathing cycles. Thirty minutes of clean breathing were randomly selected

per each infant and displayed by another custom-made Matlab tool, that allowed the clinician to visually annotate inspiration and exhalation time intervals. This procedure provided 7234 correctly identified cycles, equivalent to almost 2h30 of breathing.

### 2.1.2 Testing cohort

The second dataset is composed by a selection of sixteen infected (*Sepsis*) and sixteen non-infected infants (*No-sepsis*) paired by age, gender and weight criteria (see Table 1). The diagnostic of sepsis included the combination of an inflammatory response, i.e. C-reactive protein (CRP)  $> 5$  mg/l 24 h after the recording and positive blood cultures. In non-infected infants, no inflammatory response was observed, i.e. a CRP  $< 5$  mg/l 24 h after the recording and resulted in negative blood cultures.

Given the well-known dependence on maturation, comparing age-equivalent sick and healthy infants is mandatory in the investigation of septicemic processes [13].

	<i>Sepsis</i>	<i>No-sepsis</i>
Infants	16	16
PMA (weeks)	$30.5 \pm 1.73$	$30.4 \pm 1.64$
Postnatal age (days)	$15.6 \pm 12.2$	$15.8 \pm 10.7$
Weight (kg)	$1.11 \pm 0.27$	$1.12 \pm 0.23$
Recording time (h)	$2.14 \pm 1.07$	$2.61 \pm 0.64$

**Table 1:** Description of the testing cohort. There are no significant differences between the age and weight of infants and duration of records.

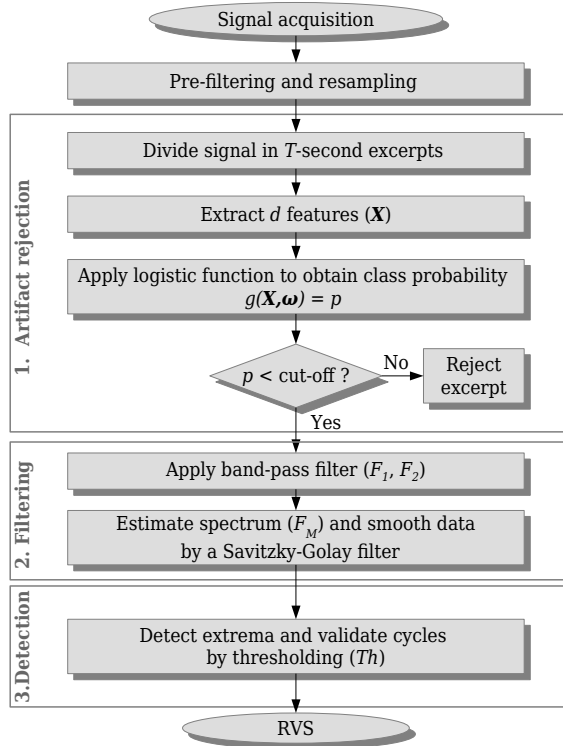
## 2.2 Methods

The processing framework to obtain RVS is composed by three blocks (see Figure 1): *i*) rejection of artifacted epochs involving gross body movements in raw signals by an automatic classifier based on logistic regression, *ii*) two-step filtering process, including band-pass and smoothing filters and *iii*) detection of the breath intervals in the clean data.

### 2.2.1 Artifact rejection

The study of the statistical distribution of the energy or root-mean square (RMS) in breathing signals is a common artifact detection criterion because in general, gross body movements induce higher amplitudes on the strain gauges. For instance, Motto et al. [14] applied this feature in breathing traces (both abdominal and thoracic) from 45 weeks PMA full term infants, employing a thresholding detector optimized by a Neyman-Pearson approach [15] that attained 89% for sensitivity and 88% for specificity. However, an energy-based threshold could be in some cases too restrictive due to the effect of deep breaths and impedance changes in the amplitudes.

On the other hand, an artifacted component could account for an unexpected transient event or for a background activity, like muscle activity or noise. Thus, in view of the noisy



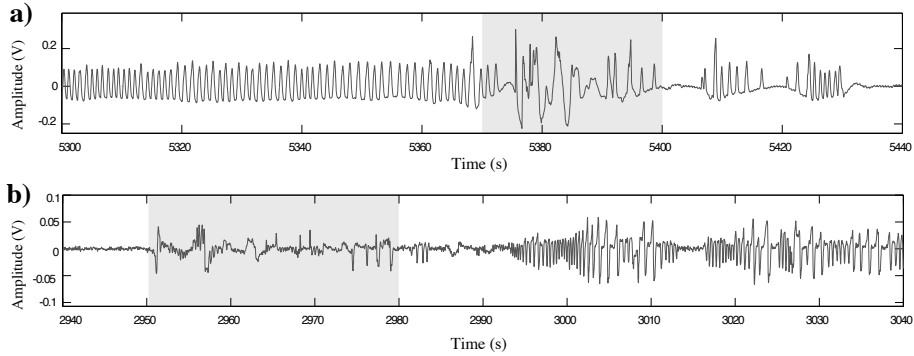
**Figure 1:** Flowchart of the proposed method to obtain clean respiratory variability series, identifying the three main blocks. 1) Artifact rejection: After segmenting the pre-processed raw signals, a set of  $d$  features ( $\mathbf{X}$ ) are obtained. A logistic function is then computed using  $\mathbf{X}$  and a vector of regression coefficients ( $\mathbf{w}$ ), learned from the gold standard. The class probability  $p$  serves to reject the breathing excerpt if this exceeds the cut-off value, chosen according to a sensitivity/specificity pair. 2) Filtering: Artifact-free data is next filtered by a band-pass filtered and its power spectrum is estimated to find the main frequency, ( $F_m$ ), necessary to find the parameters of the smoothing (Savitzky-Golay) filter. 3) Cycle detection: A peak detector governed by a threshold  $Th$  finds minima and maxima in the trace, i.e. the time instants of breaths that determine the RVS.

environments our breathing signals come from, an alternative criterion to detect the artifacts could be to measure the randomness of the traces by means of the entropy, as Mammone et al. [16] did in EEG signals by means of ICA and Renyi’s entropy. Nevertheless, the erratic breathing patterns typical in preterm infants (see Figure 2) could be an inconvenient in entropy measures and lead the classifier to false positive detections.

Therefore, a binary multi-variable model dealing with several breathing features is necessary. Among the numerous existing solutions, such as support vector machines or neural networks, we propose to employ a classifier based on logistic regression (LR) to discriminate between clean and noisy excerpts. LR is a straight-forward solution (see the appendix) providing good results in many biomedical data classification tasks [17].

From the raw breathing signal, LR exploits a set of  $d = 14$  features  $\mathbf{X} \in \mathbb{R}^d$ , based both in time and frequency domains. These features are computed on an excerpt of duration  $T$ , that we subsequently call *resp* and correspond to:

- Time-domain: The absolute difference between the maximum and minimum value ( $Mm$ ),



**Figure 2:** Unprocessed respiratory signals with artifacts (shadow zones) marked by a clinician. a) Example of a regular pattern turning into periodic after the artifacts. b) Example of a heavily contaminated signal exhibiting erratic pattern. In view of the similarities between artifacts in regular patterns and erratic patterns, the main challenge of our detector is to discriminate correctly the artifacts.

the mean of  $resp$  divided by its maximum value ( $\overline{resp}/M$ ), the standard deviation ( $Sd$ ), the kurtosis ( $Kt$ ), the maximum value of the first derivate of the excerpt ( $D1$ ), its kurtosis ( $KD1$ ) and the root mean squared value ( $RMS$ ).

- Frequency-domain: the kurtosis of the power spectral density ( $KPSD$ ) and the power computed in four different frequency bandwidths  $P_{B0}$ ,  $P_{B1}$ ,  $P_{B2}$  and  $P_{B3}$ . We also computed the main frequency ( $Mf$ ) over a period  $T$  as:

$$Mf = \frac{1}{T} \sum_{k=1}^T |resp(k) - resp(k-1)|. \quad (1)$$

- Entropy: The Shannon entropy ( $SEnt$ ) in each excerpt computed as:

$$SEnt = - \sum_{k=1}^T resp^2 \log(resp^2). \quad (2)$$

Other features were included to the model in a preliminary test (second derivative properties, number of zero-crossing, different frequency bands, etc.) but they did not improve significantly the model fit.

### 2.2.2 Band-pass and smoothing filters

As introduced before, respiratory signals in preterm infants breathing spontaneously have several peculiarities. Moreover, the notable influence of the technical artifacts present in the NICU, changes in body position and physiological events as apnea, sighs and swallows, increase the difficulty of determining appropriate filters and performing effective breath detections.

Hence, after rejecting the artifacts due to gross movements, the next step is to remove the baseline and the high frequency noise. This can be performed by a simple band-pass filter, a

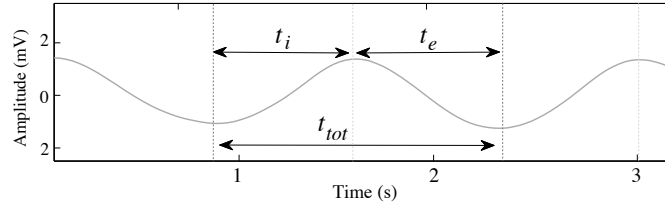
fourth-order Butterworth whose lower ( $F_1$ ) and higher ( $F_2$ ) cut-off frequencies were set respectively to 0.5 and 20 Hz. The latter was chosen to be high enough to preserve the shape of the breathing trace but not insomuch to avoid the presence of high frequency noise.

The second filter is used to smooth without distortion the signal so that cycle detection can be performed effectively. Smoothing is necessary to reduce small ripple due, for instance, to cardiac artifacts and other noise unfiltered in the previous step. This can be achieved by Savitzky-Golay (SG) filters [18], specially recommended because they preserve the width and height of peaks of the original signal, which are usually flattened by classical moving average or FIR filters.

### 2.2.3 Respiratory cycle detection

Once the breathing signal has been cleaned and smoothed, the respiratory cycles (breaths) are recognized by an automatic detector. The duration of each breath is then expressed as an univariate, cycle dependent signal, i.e. the respiratory variability series.

The proposed breath-recognition algorithm takes advantage of the sinusoidal shape of the ventilatory profile to detect the beginning and end of individual breaths. Since a minimum corresponds to the minimal wall distention, it is directly related to the end of exhalation. Then, it is considered as the start-point of a respiratory cycle. Likewise, a maximum in the respiratory trace is reached when the lungs contain the tidal volume, so the time elapsed between minima and maxima defines the inspiration time ( $t_i$ ), and conversely, the time between maxima and minima yields the exhalation time,  $t_e$  (see Figure 3).



**Figure 3:** Illustration of  $t_i$ ,  $t_e$  and  $t_{tot}$  times in a respiratory trace. Vertical lines constitute detected minima and maxima.

Detection algorithms exploiting large slopes or high frequent content in signals (for example, the QRS complex in ECG) employ transformations such as the signal derivative or the dyadic wavelet transform, producing a feature in which peaks can be easily detected by thresholding [19, 20]. However, breathing signals are sinusoidal-like, non-impulsive signals where most of the spectral power is located in a lower frequency range, typically from 0.5 to 2 Hz in newborns. Consequently, a simple extrema detector seems to be a better solution than the mentioned strategies.

Thus, the proposed detection algorithm finds all extrema, that will be validated only if the relative difference between surrounding peaks exceeds a predefined threshold,  $th$ . The pseudo-code is described in the following lines:

```

Inputs
x := Breathing signal
th := Detection threshold
Outputs
m := Detected minima
M := Detected maxima
Begin
  i_M := Find the indexes of x such that:
    x[i-1] <= x[i] >= x[i+1]
  M := Find x[i_M] such that x[i_M]-th > 0
  i_m := Find the indexes of x such that:
    x[i-1] >= x[i] <= x[i+1]
  m := Find x[i_m] such that x[i_m]+th > 0
End.

```

The detection threshold  $th$  is obtained by the product of a descriptive statistic of the respiratory signal and a constant  $\kappa$ . The employed statistic was the inter-quartile range (IQR) of the breathing amplitudes because it constitutes a measure of dispersion excluding the 25% of extrema, so deep and shallow breaths are not taken into account in the detection.

### 3 Results

In this section, the artifact rejection and cycle detection are validated employing the two gold standards (5161 ten-second excerpts classified as clean/artifacted and 7234 correctly identified cycles) introduced in Section 2.1.1. Their performances, as well as the parameter tuning of the filtering module, are also described.

#### 3.1 Automatic artifact rejection

Before computing the LR model, we first identified experimentally the bandwidths  $B_0$ ,  $B_1$ ,  $B_2$  and  $B_3$  where the powers  $P_{B0}$ ,  $P_{B1}$ ,  $P_{B2}$  and  $P_{B3}$  were maximum. The following frequency intervals were found:  $B_0 = [0, 0.25]$ ,  $B_1 = [0.25, 0.7]$ ,  $B_2 = [0.7, 1.2]$  and  $B_3 = [1.2, 2]Hz$ . Next, employing the 14 features –computed in sliding excerpts of  $T = 10$  s– on the five infants, a logistic function was found by a maximum likelihood estimate (MLE) using the *glmfit* function in the Statistical Toolbox of Matlab. Finally, a Wald test (see appendix) was performed to retain the most contributive features. This resulted in reducing the initial 14 variables to seven (detailed in Table 2). The remaining features did not contribute significantly to the model since they failed to reject the null hypothesis ( $p > 0.05$ ).

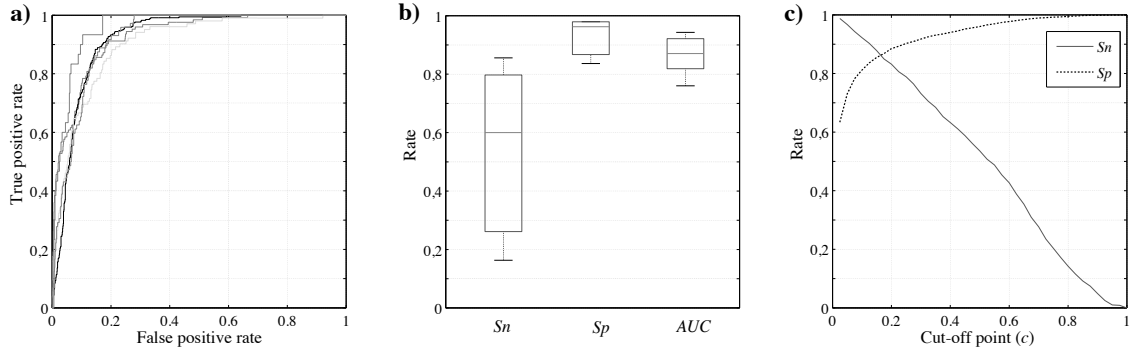
Dep var →	Intercept	$Mm$	$RMS$	$MF$	$KPSD$	$P_{B0}$	$P_{B3}$	$SEnt$
$\hat{w}$	1.865	-0.743	0.152	-0.009	0.968	0.030	-5.863	-7.176
$SE$	0.579	0.144	0.069	0.001	0.262	0.008	0.652	0.762
$z$	3.220	5.184	2.192	5.856	3.687	3.683	8.987	9.407
$p$	< 0.01	< 0.001	0.028	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

**Table 2:** Results from fitting the LR model to the artifact detector. The estimated coefficients ( $\hat{w}$ ) corresponding to the seven selected features plus the intercept, and the standard errors ( $SE$ ) of the MLE are detailed. We also provide the  $z$ -statistics and  $p$ -values of the Wald test.

To see the efficacy of the artifact classifiers, receiver operating characteristic (ROC) curves (see Figure 4a) are plotted using leave-one out cross-validation (LOOCV). In Figure 4b) sensi-

bilities, sensitivities and areas under curve are obtained using a logit cut-off value of 0.5, but to find a cost-effective point, we represent the mean of both measures as a function of  $c$  (Figure 4c) and select the intersection point ( $c = 0.165$ ), that corresponds to a 86% sensitivity ( $Sn$ ) and 86% specificity ( $Sp$ ).

The less optimal values obtained with this detector ( $Sn = 86\%$ - $Sp = 86\%$  against  $Sn = 89\%$ - $Sp = 89\%$  in full-terms [14]) are probably due to the complexity of the classification problem in more immature breathing patterns. Indeed, the less performing ROC curve corresponds to a patient having patterns predominantly erratic (a sample can be seen in Figure 2b) whereas the best one contains more regular patterns.



**Figure 4:** a) ROC curves from LOOCV. b) Sensitivity ( $Sn$ ) and specificity ( $Sp$ ) based on the default cut-off value ( $c = 0.5$ ). Areas under curve ( $AUC$ ) from ROC curves are also shown as a measure of performance. c) Mean of the sensitivity/specificity pair at several cut-off values.

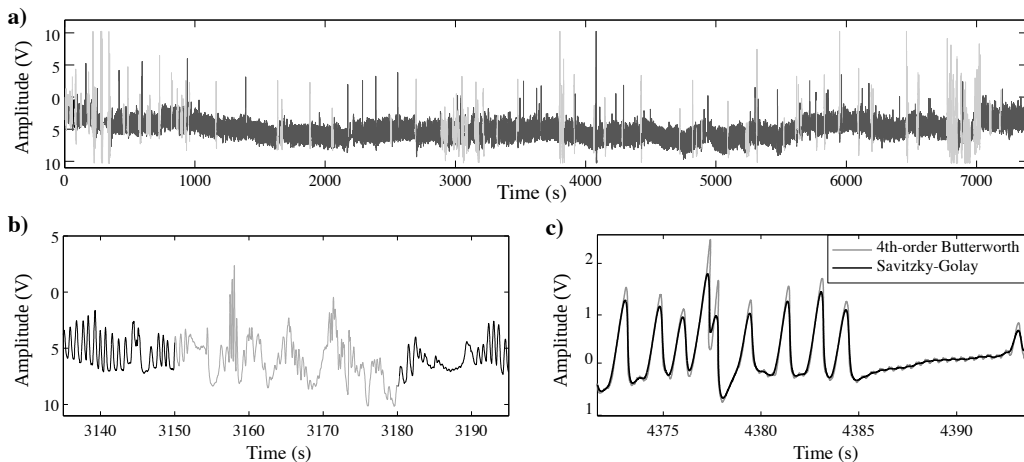
### 3.2 Band-pass and smoothing filter

The resulting signal, ignoring the artifact period, is then filtered by the band-pass and Savitzky-Golay filters. It is also used to estimate the main frequency  $F_M$  in the power spectrum. This information, necessary to tune the second filter, was estimated by means of a 30th order Burg autoregressive model.

In SG filters, data is smoothed by applying local least-squares polynomial approximation. Applied to oversampled signals (all signals in the NICU are sampled at 400 Hz whereas respiratory oscillations are mainly about 1 Hz) corrupted by noise, SG filters match appropriately the original waveform with fitted polynomial slopes reducing high-frequency noise [21].

Two parameters need to be chosen to apply this filter: the polynomial order ( $n$ ) and the length of frame ( $m$ ). To perform an optimal smoothing, it should be considered that  $n = 2$  takes into account the curvature of the signals and  $n = 3$  the inflection points. Larger values are unnecessary and  $n = 0$  produces the same effect than a moving average filter. Generally,  $m$  needs to be larger than  $n$ . If  $m \simeq n$  the interpolation uses almost as many points as those contained in the original signal and no smoothing is produced.

After some tests, we found experimentally that an optimal filtering was obtained by fixing  $n = 2$  and  $m = \frac{F_s}{2F_M}$ .



**Figure 5:** a) Example of artifact detection (excerpts in gray) in a patient not used in the training set. As the baseline evidences, no filtering is still performed at this step. b) Detail of detected artifacts. c) Detail of the signal, after being processed by the two filters. Note that the small ripple disappears after the SG filter.

An example of a raw signal, first processed by the artifact detector and then by the filters, is shown in Figure 5.

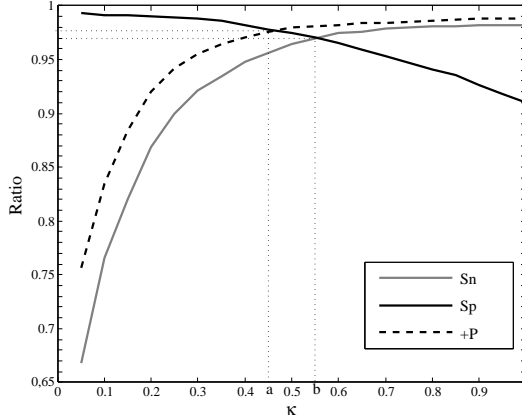
### 3.3 Performance of the cycle detector

Following the guidelines proposed by the Association for the Advancement of Medical Instrumentation [24], the parameters used to evaluate the performance of the breath detection algorithm were based on the traditionally-employed measures for heart beat detectors. These were sensitivity and positive predictive value,  $+P$ , based upon the measures of true positives ( $TP$ ) and false positives ( $FP$ ) by the expression  $+P = TP/(TP + FP)$ .

The automatic detector was applied to the dataset having the cycles identified by the clinician, varying the threshold  $th$  multiplied by a coefficient  $\kappa$  from 0.05 to 1 in steps of 0.05. Since the measures of performance considered the number of correctly detected extrema, a peak was classified as valid if there was a match within the surrounding 10% of the cycle time with a manual label. Otherwise, it was counted as false positive. Figure 6 shows the results of the performances for thresholds based on the IQR, attaining  $Sn = Sp = 97\%$  or  $Sp = +P = 97.7\%$  depending on the cut-off point.

## 4 Clinical application: detection of infection

As the diagnosis of sepsis still needs blood culture – invasive, slow and low predictive in the earliest phases of infection [23] –, in last years it has been an increasing interest on alternative tools based on the analysis of biomedical signals. In this section we propose a linear analysis, employing standard statistical measures and a basic spectrum analysis, to assess the potential of RVS to detect sepsis.



**Figure 6:** Performance plot of the cycle detector, applying the  $\kappa$  coefficient to the IQR of the breathing amplitudes. Two possible optimal operating points are indicated: a) The crossing of specificity ( $Sp$ ) and positive predictive value ( $+P$ ) in  $\kappa=0.46$  and b) the crossing of specificity and sensitivity in  $\kappa=0.55$ .

#### 4.1 Selection of variables

The testing cohort was firstly preprocessed by the noise rejection algorithm, the filters and the cycle detection module. Clean respiratory variability time series  $t_{tot}$ ,  $t_i$ ,  $t_e$  were then built to find differences between *Sepsis* and *No-sepsis* groups. A first comparison of the general descriptors of breathing could be done between groups. These descriptors reflect the temporal distribution and irregularity of breathing:

- Concerning cycles: The mean duration of cycle times ( $\overline{t_{tot}}$ ) in seconds and their standard deviations ( $\sigma_{t_{tot}}$ ) to describe breathing variability were estimated. To describe the shape of breathing cycles, we computed the mean ratio of inspiration and exhalation times  $t_i/t_e$ .
- Concerning apnea: The average number of apneas (it is defined as a cessation of the ventilatory activity during at least 15 seconds), normalized in one hour ( $N_{ap}$ ), their mean duration ( $D_{ap}$ ) and their standard deviation ( $\sigma_{ap}$ ) in seconds were computed.

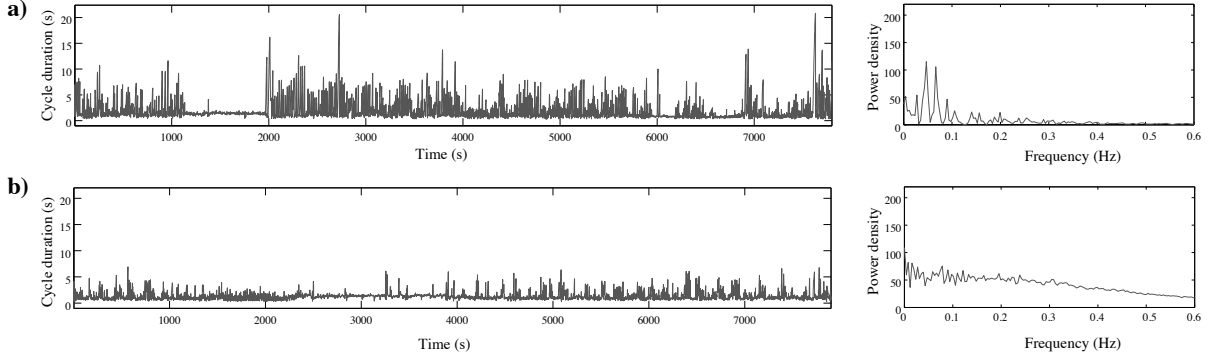
Inspired by the spectral analysis of the heart beat ratio, we hypothesize that some system's control information may be obtained by investigating periodic components in the respiratory rhythm. In RVS spectrum, only data about the duration of cycles is processed and bias due to amplitude variations in breaths can be avoided. Indeed, temporal signals are directly measured from abdominal movements –uncalibrated and noisy– and a spectral analysis would not provide reliable information.

Therefore, we computed the power spectra using the Fast Fourier Transform (FFT) on RVS (see Figure 7). To this end, the series were previously resampled-interpolated to obtain equidistantly time sampled signals at 2 Hz, as done traditionally with heart rate variability series [22], dividing the range of frequencies into four components expressed as the percentage relative to the total spectral power:

- VLF, or very low frequency: defined from 0 to 0.05 Hz.

- LF, or low frequency: from 0.05 to 0.15 Hz.
- MF, or medium frequency: from 0.15 to 0.25 Hz.
- HF, or high frequency: from 0.25 to 0.60 Hz.

Note that the above frequencies have been chosen arbitrarily and the components are not equivalent to those typically studied for heart rate variability because of the different dynamics of breathing.



**Figure 7:** (a) RVS having periodic breathing and its power density. Periodicities can be found in the spectrum between 0.05 and 0.1 Hz. (b) RVS predominantly regular, characterized by a more flat spectrum.

## 4.2 Results and discussion

The results comparing the respiratory variables between healthy and infected patients are summarized in Table 3. They show that *Sepsis* patients have in average slightly longer  $t_{tot}$ , a higher number of apneas  $N_{ap}$ , higher VLF and lower LF contents, although statistical tests did not find significance.

	<i>Sepsis</i>	<i>No-sepsis</i>
$t_{tot}$ (s)	$1.49 \pm 0.45$	$1.37 \pm 0.26$
$\sigma t_{tot}$	<b><math>1.27 \pm 0.59</math></b>	<b><math>0.84 \pm 0.24^*</math></b>
$t_i/t_e$	$1.17 \pm 0.10$	$1.12 \pm 0.07$
$N_{ap}$	$50.6 \pm 33.7$	$41.1 \pm 34.0$
$D_{ap}$ (s)	<b><math>8.09 \pm 2.84</math></b>	<b><math>5.91 \pm 1.20^*</math></b>
$\sigma_{ap}$	<b><math>3.48 \pm 2.06</math></b>	<b><math>2.01 \pm 1.01^*</math></b>
VLF (%)	$66.30 \pm 19.93$	$56.87 \pm 19.86$
LF (%)	$29.98 \pm 17.20$	$39.15 \pm 17.52$
HF (%)	$3.32 \pm 3.43$	$3.77 \pm 2.85$

**Table 3:** Results from the comparative study of *Sepsis* and *No-sepsis* populations. The variables are expressed as mean  $\pm$  std. dev. Asterisks denote statistically significant differences in a Mann-Whitney U test ( $p < 0.05$ ).

On the other hand, the standard deviation of the cycle duration ( $\sigma t_{tot}$ ) was significantly higher, a fact that reveals more unstable breathing patterns in sick infants. Moreover, the mean duration of apneas,  $D_{ap}$ , as its standard deviation,  $\sigma_{ap}$ , are substantially longer.

Therefore, it can be stated that breathing in infected infants is, in general, more variable concerning the duration of breaths and characterized by longer episodes of apnea. Considering the relationships between apnea and bradycardia, these results are complementary to the findings related to cardiac variability, in which severe, unusual and recurrent bradycardias are documented in infected populations [11, 25].

## 5 Conclusion

This paper proposes a new framework to process the raw breathing signals acquired in the NICU to obtain clean respiratory traces and their respective respiratory variability series. Despite the probability of including corrupted segments (or excluding good segments) is non-negligible, the detector can be adjusted by modifying the cut-off point  $c$  in the logistic regression classifier. If the subsequent application does not tolerate noise, such as certain nonlinear analyses of the temporal signal,  $c$  should maximize sensitivity at the cost of losing useful information. Contrarily, applications more robust to noise and needing long records, could deal with lower sensitivities.

Because the normal development of respiratory controlling systems is associated with characteristic changes in linear and nonlinear measures, we proposed here the statistical characteristics of the breathing cycles, apneas and the spectra of RVS in relation with the infection. As expected, most of the measures obtained from our data (duration of breaths, apneas) confirmed their evolution in accordance to infection, evidencing functional and anatomical changes in the respiratory system. Indeed, linear parameters evidenced the existence of more irregular breathing cycles in sick infants, as well as longer apneas.

In conclusion, on our data sets, results were consistent to other reference works and should be useful to quantify preterm infants' maturity and the proposed detector becomes a powerful tool to obtain respiratory variability series, that can be analyzed by linear tools or nonlinear methods such as we already reported in [12].

## Acknowledgment

The authors would like to thank the clinicians from the Pôle Médico-Chirurgical de Pédiatrie et de Génétique Clinique of the CHU of Rennes for their large involvement in this study.

## Appendix: Logistic regression-based classifiers

Logistic regression (LR) is a type of regression analysis derived from the linear case. In linear regression, a function  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  related to the probabilities  $\pi_i$ , is expressed as a linear combination of  $\mathbf{X}_i$ :

$$g(\mathbf{X}_i) = w_0 + w_1 X_{1,i}, \dots, w_d X_{d,i} = w_0 + \mathbf{w} \cdot \mathbf{X}_i, \quad (3)$$

where  $\mathbf{w} = [w_1, \dots, w_d]$  is the vector of regression coefficients and  $w_0$  is the intercept coefficient. In logistic regression,  $\pi_i$  is expressed through a function called logit, of the form:

$$\text{logit}(\pi_i) = \ln \left( \frac{\pi_i}{1 - \pi_i} \right) = w_0 + \mathbf{w} \cdot \mathbf{X}_i. \quad (4)$$

The inverse of the above expression, called logistic function, is expressed as:

$$\text{logit}^{-1}(\pi_i) = \frac{1}{1 + e^{-(w_0 + \mathbf{w} \cdot \mathbf{X}_i)}} = g(\mathbf{X}_i, \mathbf{w}). \quad (5)$$

An important characteristic of the logistic function is that it is bounded between 0 and 1, and thus, it can be used directly to model the probabilities of the possible outcomes:

$$P(Y = 1 | \mathbf{w}, \mathbf{X}_i) = g(\mathbf{X}_i, \mathbf{w}). \quad (6)$$

Given the set of  $N$  examples  $\{\mathbf{X}_i, Y_i\}$ , the learning process aims at finding the best  $\mathbf{w}$ , which is to maximize the conditional probabilities  $P(Y_i | \mathbf{X}_i, \mathbf{w})$ . This can be achieved by maximizing the log likelihood function:

$$L(\mathbf{w}) = \sum_{i=1}^N Y_i \ln P(Y_i = 1 | \mathbf{X}_i, \mathbf{w}) + (1 - Y_i) \ln P(Y_i = 0 | \mathbf{X}_i, \mathbf{w}). \quad (7)$$

The maximum likelihood estimate (MLE) gives the optimal coefficients,  $\hat{\mathbf{w}}$ , employed to estimate the probabilities,  $\hat{\pi}_i$ . It is found by a numerical optimization, a process that starts with a set of initial values and iterate according to a hill-climbing algorithm such as Newton-Raphson's method.

Several approaches exist to assess the goodness of the fitted model. The principle of some of them is to estimate two models and compare the fits removing predictor variables from one of the two models. By doing this, the model having less variables (the most restrictive) fits less well, showing a lower log likelihood. The likelihood ratio test, for instance, compares the log likelihoods of the two models and if this difference is statistically significant, then the less restrictive model is chosen as it fits the data significantly better.

The Wald test approximates the likelihood ratio test but only requires estimating one model. Essentially, it tests the null hypothesis that a set of parameters is equal to some value, zero if some variable is removed. If the test fails to reject the null hypothesis, then removing the variables from the model does not harm the fit of that model significantly, because a predictor with a coefficient that is very small relative to its standard error does not generally contribute substantially predict the dependent variable. The contribution of a given coefficient  $\hat{w}_j$  to the model can be quantified by the  $z$ -statistic:

$$z = \frac{\hat{w}_j}{SE}, \quad (8)$$

where  $SE$  is the standard error (an estimation of the standard deviation) of the estimated parameter.

The decision of class membership given by the activation function is bounded by a cut-off value  $c$ , such that  $f(\hat{\pi}_i) > c$  assigns the predictive output value,  $\hat{y}$ , to class 1, and  $f(\hat{\pi}_i) \leq c$  assigns  $\hat{y}$  to class 0. By default  $c = 0.5$ , but it can be adjusted to certain specificity or sensitivity specifications.

## References

- [1] H. C. Miller, F. C. Behrle, and N. W. Smull, "Severe apnea and irregular respiratory rhythms among premature infants: A clinical and laboratory study," *Pediatrics*, vol. 23, no. 4, pp. 676–685, Apr. 1959.
- [2] E. R. Alden, T. Mandelkorn, D. E. Woodrum, R. P. Wennberg, C. R. Parks, and W. A. Hodson, "Morbidity and mortality of infants weighing less than 1,000 grams in an intensive care nursery," *Pediatrics*, vol. 50, no. 1, pp. 40–49, Jul. 1972.
- [3] M. Durand, L. A. Cabal, F. Gonzalez, S. Georgie, C. Barberis, T. Hoppenbrouwers, and J. E. Hodgman, "Ventilatory control and carbon dioxide response in preterm infants with idiopathic apnea," *American journal of diseases of children (1960)*, vol. 139, no. 7, pp. 717–720, Jul. 1985.
- [4] R. J. Martin, M. J. Miller, and W. A. Carlo, "Pathogenesis of apnea in preterm infants," *The Journal of pediatrics*, vol. 109, no. 5, pp. 733–741, Nov. 1986.
- [5] C. F. Poets, V. A. Stebbens, M. P. Samuels, and D. P. Southall, "The relationship between bradycardia, apnea, and hypoxemia in preterm infants," *Pediatric research*, vol. 34, no. 2, pp. 144–147, Aug. 1993.
- [6] H. Rigatto and J. P. Brady, "Periodic breathing and apnea in preterm infants. i. evidence for hypoventilation possibly due to central respiratory depression," *Pediatrics*, vol. 50, no. 2, pp. 202–218, Aug. 1972.
- [7] G. Pichler, B. Urlesberger, and W. M. ller, "Impact of bradycardia on cerebral oxygenation and cerebral blood volume during apnoea in preterm infants," *Physiological Measurement*, vol. 24, no. 3, pp. 671–680, Aug. 2003.
- [8] A. Janvier, M. Khairy, A. Kokkotis, C. Cormier, D. Messmer, and K. J. Barrington, "Apnea is associated with neurodevelopmental impairment in very low birth weight infants," *Journal of perinatology: official journal of the California Perinatal Association*, vol. 24, no. 12, pp. 763–768, Dec. 2004.
- [9] M. Altuve, G. Carrault, J. Cruz, A. Beuchee, P. Pladys, and A. Hernandez, "Analysis of the QRS complex for apnea-bradycardia characterization in preterm infants," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009. EMBC 2009*. IEEE, Sep. 2009, pp. 946–949.
- [10] H. Cao, D. E. Lake, M. P. Griffin, and J. R. Moorman, "Increased nonstationarity of neonatal heart rate before the clinical diagnosis of sepsis," *Annals of Biomedical Engineering*, vol. 32, no. 2, pp. 233–244, Feb. 2004.
- [11] A. Beuchée, G. Carrault, J. Y. Bansard, E. Boutaric, P. Bétrémieux, and P. Pladys, "Uncorrelated randomness of the heart rate is associated with sepsis in sick premature infants," *Neonatology*, vol. 96, no. 2, pp. 109–114, 2009.
- [12] X. Navarro, F. Porée, A. Beuchée, and G. Carrault, "Performance analysis of hurst exponent estimators using surrogate-data and fractional lognormal noise models: Application to breathing signals from preterm infants," *Digital Signal Processing*, vol. 23, no. 5, pp. 1610–1619, 2013.
- [13] R. A. Polin, "The ins and outs of neonatal sepsis," *Journal of Pediatrics*, vol. 143, no. 1, pp. 3–4, 2003.
- [14] A. L. Motto, H. L. Galiana, K. A. Brown, and R. E. Kearney, "Detection of movement artifacts in respiratory inductance plethysmography performance analysis of a neyman pearson energy based detector," *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2004, IEMBS'04*, vol. 1, pp. 49–52, Sep. 2004.
- [15] S. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*, 1st ed. Prentice Hall, Feb. 1998.
- [16] N. Mammone and F. C. Morabito, "Enhanced automatic artifact detection based on independent component analysis and Renyi's entropy," *Neural Networks*, vol. 21, no. 7, pp. 1029–1040, Sep. 2008.
- [17] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of Biomedical Informatics*, vol. 35, no. 5, pp. 352–359, 2002.
- [18] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, pp. 1627–1639, 1964.
- [19] W. P. Holsinger, K. M. Kempner, and M. H. Miller, "A QRS preprocessor based on digital differentiation," *IEEE Transactions on Biomedical Engineering*, vol. BME-18, no. 3, pp. 212–217, May 1971.

- [20] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230–236, Mar. 1985.
- [21] R. Schafer, "What is a Savitzky-Golay filter? [lecture notes]," *IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 111–117, Jul. 2011.
- [22] M. Malik, J. T. Bigger, A. J. Camm, R.E. Kleiger, A. Malliani, A.J. Moss and P.J. Schwartz, "Heart rate variability standards of measurement, physiological interpretation, and clinical use". *European heart journal*, vol. 17(3), pp. 354–381, 1996.
- [23] A. Malik, C. P. S. Hui, R. A. Pennie, and H. Kirpalani, "Beyond the complete blood cell count and c-reactive protein: a systematic review of modern diagnostic tests for neonatal sepsis," *Archives of Pediatrics & Adolescent Medicine*, vol. 157, no. 6, pp. 511–516, Jun. 2003.
- [24] ANSI/AAMI CE57, "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms. Recommended Practice/American national standard. <http://www.aami.org>," 1998.
- [25] M. P. Griffin, T. M. O’Shea, E. A. Bissonette, F. E. Harrell, D. E. Lake, and J. R. Moorman, "Abnormal heart rate characteristics preceding neonatal sepsis and sepsis-like illness," *Pediatric Research*, vol. 53, no. 6, pp. 920–926, 2003.