

# BreathFinder: A Method for Non-Invasive Isolation of Respiratory Cycles Utilizing the Thoracic Respiratory Inductance Plethysmography Signal

Benedikt Holm<sup>1</sup>, Michal Borsky<sup>1</sup>, Erna S Arnardottir<sup>2,3</sup>, Marta Serwatko<sup>2</sup>, Jacky Mallett<sup>1</sup>, Anna Sigridur Islind<sup>1</sup>, María Óskarsdóttir<sup>1</sup>

<sup>1</sup>Reykjavik University, School of Technology, Department of Computer Science, Reykjavik, Iceland; <sup>2</sup>Reykjavik University, School of Technology, Sleep Institute, Reykjavik, Iceland; <sup>3</sup>Landspítali, The National University Hospital of Iceland, Reykjavik, Iceland

Correspondence: Benedikt Holm, Email benedikhth@ru.is

**Introduction:** The field of automatic respiratory analysis focuses mainly on breath detection on signals such as audio recordings, or nasal flow measurement, which suffer from issues with background noise and other disturbances. Here we introduce a novel algorithm designed to isolate individual respiratory cycles on a thoracic respiratory inductance plethysmography signal using the non-invasive signal of the respiratory inductance plethysmography belts.

**Purpose:** The algorithm locates breaths using signal processing and statistical methods on the thoracic respiratory inductance plethysmography belt and enables the analysis of sleep data on an individual breath level.

**Patients and Methods:** The algorithm was evaluated against a cohort of 31 participants, both healthy and diagnosed with obstructive sleep apnea. The dataset consisted of 13 female and 18 male participants between the ages of 20 and 69. The algorithm was evaluated on 7.3 hours of hand-annotated data from the cohort, or 8782 individual breaths in total. The algorithm was specifically evaluated on a dataset containing many sleep-disordered breathing events to confirm that it did not suffer in terms of accuracy when detecting breaths in the presence of sleep-disordered breathing. The algorithm was also evaluated across many participants, and we found that its accuracy was consistent across people. Source code for the algorithm was made public via an open-source Python library.

**Results:** The proposed algorithm achieved an estimated 94% accuracy when detecting breaths in respiratory signals while producing false positives that amount to only 5% of the total number of detections. The accuracy was not affected by the presence of respiratory related events, such as obstructive apneas or snoring.

**Conclusion:** This work presents an automatic respiratory cycle algorithm suitable for use as an analytical tool for research based on individual breaths in sleep recordings that include respiratory inductance plethysmography.

**Keywords:** respiratory analysis, breath detection algorithm, sleep analysis, breath segmentation, respiratory cycle isolation

## Introduction

At present, to be able to correctly diagnose a sleep disorder, an expert sleep scorer must manually review (score) a Polysomnography (PSG) which is an overnight collection of various physiological signals from a patient suffering from a suspected sleep disorder. This type of study is performed either in a controlled hospital environment or in a home setting, each with their advantages and disadvantages.<sup>1</sup> A wide range of signals is currently being collected, including respiratory inductance plethysmography (RIP), oxygen saturation (SpO<sub>2</sub>), nasal airflow, electroencephalography (EEG), electromyograms (EMG), electrocardiography (ECG), audio, and others.<sup>2</sup> The sleep scorer must annotate the sleep stages and other events of interest, which include respiratory events (apneas, hypopneas), oxygen desaturations, body movements, and respiratory event related arousals (brief waking periods due to breathing interruptions). These annotations are then used to determine the diagnosis and recommend a treatment.

For historical reasons, most automated scoring of PSG data uses fixed-length epochs following the methodology adopted for manual scoring. An alternative approach, adaptive segmentation, based on segments of variable length depending on the signal,<sup>3</sup> has been confined to research on brain activity during sleep,<sup>4</sup> with some success in sleep staging.<sup>5,6</sup>

A limited amount of literature exists on using adaptive segmentation to identify individual breaths, or respiratory cycle isolation (RCI), with existing methods mainly based on statistical analysis of signals such as peak and valley detection in the airflow, thoracic or abdominal RIP signals, and feature extraction and modeling, which are mostly derived from the audio signal recorded during the study.<sup>7–15</sup> One problem with some approaches is that formal validation of the algorithm is often not provided on patient data. Issues with equipment, wide variations in patient behavior, and noisy environments such as partner breathing or background noise, may consequently not have been adequately explored.

Moyles & Erlandson proposed a non-parametric statistical approach to RCI, based on detecting changes in the trend of the airflow signal, but do not provide any validation of their algorithm.<sup>12</sup> Korten & Haddad presented a pattern recognition algorithm that detects respiratory events in a barometric pressure signal,<sup>16</sup> they claim the difference in mean values for inspiratory time ( $T_i$ ), expiratory time ( $T_e$ ), and total respiratory cycle time ( $T_{tot}$ ) between the manually calculated values and the automatically detected values using the pattern recognition algorithm is very small (<6%), but do not explicitly state performance in terms of detections.

Lopez-Meyer et al presented an RCI algorithm based on peak and valley detection on RIP signals to determine the beginning and end of a breath segment, reporting 96% precision when detecting breath cycles for participants during rest.<sup>11</sup> A Python library, RespInPeace, for RIP belt analysis is also available, which uses a peak and valley location algorithm to find respiratory cycles during a conversation. but appears to have no published validation.<sup>14</sup>

Although the existing literature presents diverse methods for RCI algorithms, there is no consensus yet on which signals to base the segmentation on, validation of methods is limited and whether the segmentation should be done on a respiratory phase basis, or on a respiratory cycle basis is not always clear, with the task of RCI sometimes referred to as breath segmentation,<sup>11,17</sup> or breath cycle segmentation.<sup>18</sup> The problem of performing RCI on PSG data in relation to sleep and respiratory events does not appear to have been deeply studied.

In the rest of this paper, we will present and evaluate BreathFinder: a novel algorithm that locates individual respiratory cycles within breathing signals collected from a PSG using signal processing and statistical methods. This research aims to enhance the analysis of sleep data on an individual breath level. We evaluate our method on a real-life dataset of over eight thousand individual breaths. The main contributions of this paper are:

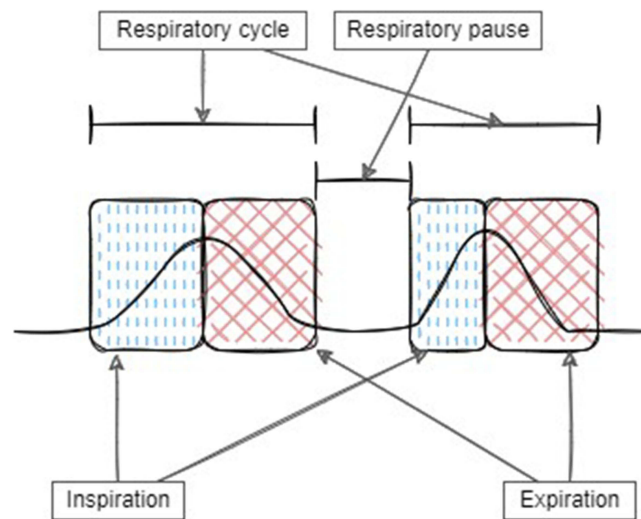
- A novel algorithm for performing RCI.
- New methods for evaluating the performance of algorithms designed to locate events in signals.

## Materials and Methods

The common definition of a respiratory cycle in the literature splits a single cycle into 4 distinct phases; inspiratory, inspiratory pause, expiratory, and expiratory pause.<sup>10</sup> In this work, a single cycle in the respiratory system is defined as starting with an inhalation and ending just after the following exhalation, with the terms “breath” and “respiratory cycle” are considered synonymous.

This definition ignores the inspiratory pause phase and interprets the expiratory pause phase as a pause between two individual breaths belonging to neither.

This definition also explicitly defines that by its definition of breath, no two breaths can occupy the same moment in time. The different phases of the respiratory cycle as defined in this work are visualized in [Figure 1](#). We used breathing signals from the PSG to detect respiratory events, particularly airflow and RIP signals. The airflow signal measures nasal respiration and is most commonly measured with a pressure transducer attached to a nasal cannula.<sup>2</sup> RIP signals are measured via two belts that stretch around the thorax and abdomen to measure changes in inductance caused by the movement of the body part in which they are placed. RIP belts are normally used to detect respiratory events with the nasal cannula and estimate respiratory effort.



**Figure 1** Phases of the respiratory cycle on the thoracic respiratory inductance plethysmography signal.

When performing RCI, a decision must be made on the signal source which is most appropriate for this purpose. The two main factors in this decision are the error rate of the signals and any potential impacts of external factors such as background noise.

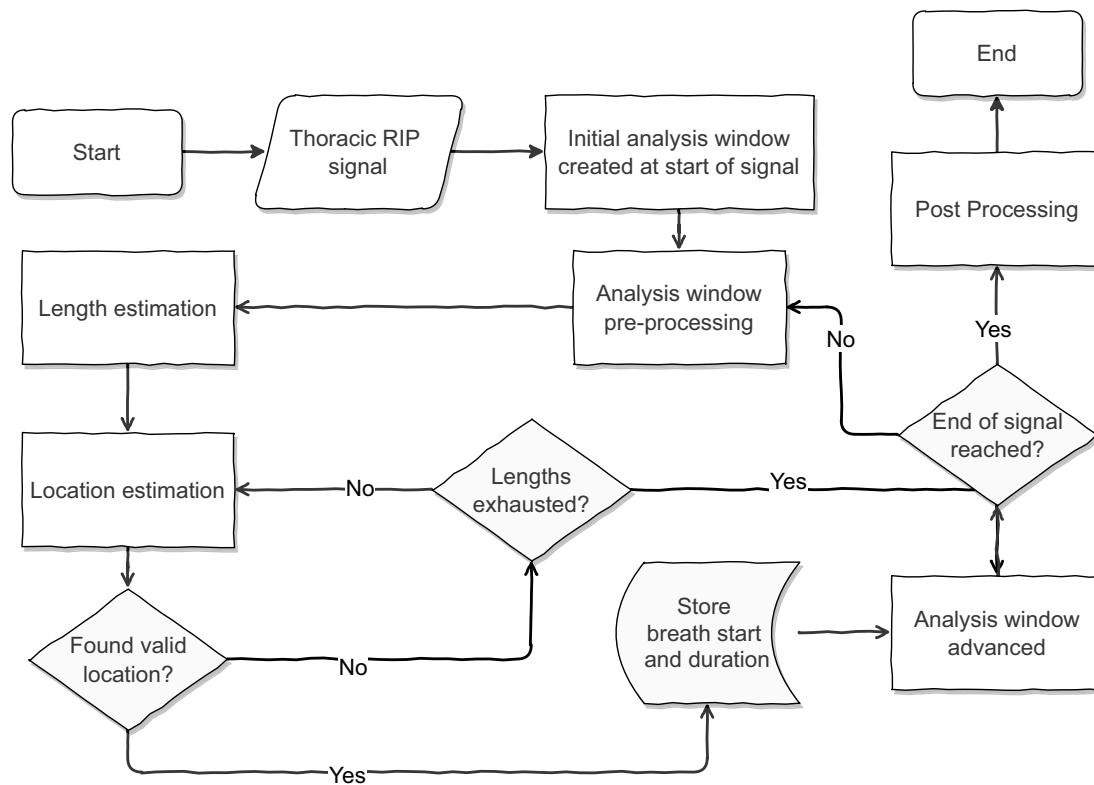
In practice the nasal cannula has several logistical issues: the sensor can get loose, affecting the measured airflow, or the participant can start mouth breathing, bypassing the sensor completely. Since multiple studies also show that the nasal airflow signal exhibits poor quality in an estimated 10% of cases,<sup>19,20</sup> we deemed it inappropriate for this study. The audio signal was also eliminated, even though some studies show the signal is not as prone to error as the other signals,<sup>20</sup> because the signal may contain many different acoustic events such as snoring, movement-related artefacts, or various background noises which complicate the task of pure breath detection.<sup>9</sup> RIP belts have the advantage that they are not susceptible to ambient noises, nor the bypass problem that the airflow signal may encounter. Of the two RIP belts, since the thoracic RIP signal captures the action of the chest-wall muscles more closely than the abdominal signal, we chose the thoracic RIP signal as the basis for our analysis.

## The BreathFinder Algorithm

A flowchart of the proposed RCI algorithm is presented in [Figure 2](#). The algorithm takes a thoracic RIP signal as a parameter, along with a sampling frequency  $f_s$ . The output is a list of individual respiratory cycles, each consisting of the onset, ie the start of the respiratory cycle in seconds since the start of the signal, and the duration of the respiratory cycle in seconds. The algorithm works on the principle of segmenting the signal into windows  $w[n]$ , with arbitrary onset  $n$  in the signal, and then searching for a single respiratory cycle in the selected window. The algorithm first calculates the autocorrelation function for  $w[n]$  to estimate the lengths  $l$  of all potential breaths in the window. It then uses a probability model to discard breath length candidates that are considered too unlikely, either because the length is too long or too short. Then, for each remaining breath length  $l$ , the algorithm creates a template waveform of that length, which correlates with the signal window to find where in the window the breath onset is most likely to be. After the window is analyzed, the algorithm advances the window further in the signal, and repeats the process. The analysis windows overlap to allow the algorithm to analyze every breath multiple times.

## Signal Pre-Processing

The preprocessing of the RIP signal is two-fold. First, the entire signal is smoothed, using a Savitzky-Golay filter, which fits a polynomial function to smooth the data points.<sup>21</sup> Here, we used a filter employed with a third order polynomial over every two seconds of signal data. The filter parameters were tuned beforehand via experimentation to ensure that the smoothing minimally affected the overall shape of the RIP signal while eliminating some of the finer noise. Then, each



**Figure 2** Respiratory Cycle Isolation algorithm flowchart.

individual  $w[n]$  was corrected for skew. This was achieved by fitting a linear function to the signal in  $w[n]$ , and adjusting the function so that the y-intercept was 0.0. The function was then subtracted from each sample of the signal. This procedure removes large-scale skew from the signal window but leaves the general shape of the signal intact. The procedure also had a positive effect on the template waveform fitting procedure, making it less likely to produce incorrect results due to skew. The result of this pre-processing step was that the cleaned thoracic RIP signal was ready to be used to estimate breath lengths.

### Main Algorithm Body

In the first step, the algorithm takes an analysis window  $w[n]$ , containing a cleaned signal and estimates its periodicity  $T$  using the autocorrelation function. The principle of the autocorrelation function (ACF) is to shift the signal forwards in time by  $k$  and to compare it to itself. When  $k = 0$ , the signal correlates perfectly with itself, but as  $k$  increases, the correlation decreases.

The formula for autocorrelation of a signal  $x$  is:

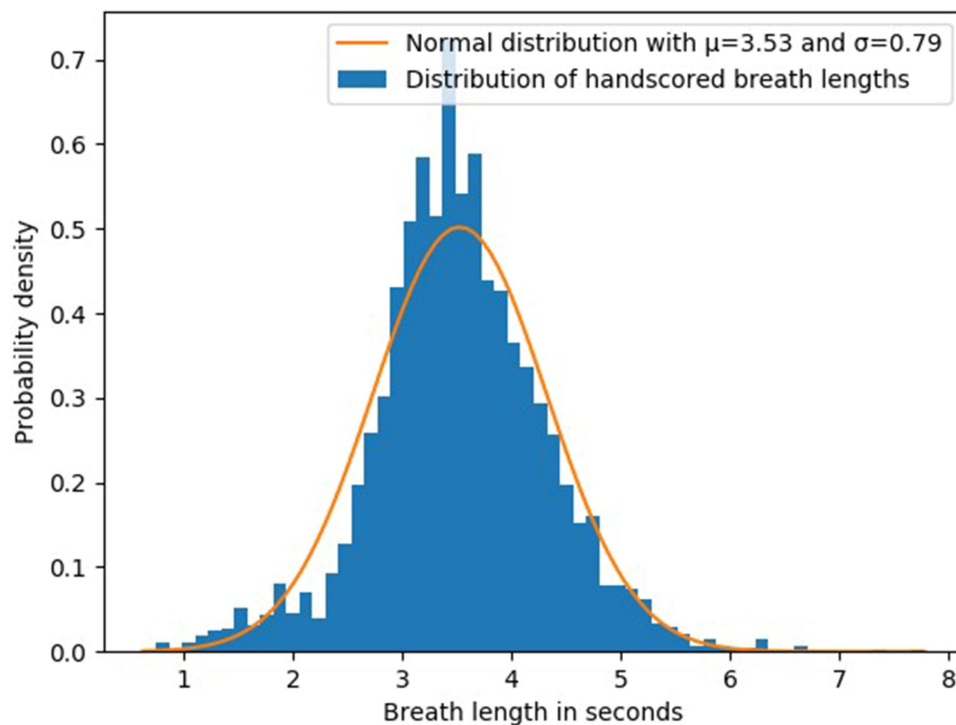
$$ACF(x)[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[k] \cdot x[n+k]$$

where  $N$  is the length of  $x$ , and  $k$  the shift.

For periodic signals, when  $k = \frac{T}{2}$ , the value of the auto correlation is low, as the signal is being compared to itself when it is in asynchrony. As  $k$  approaches  $T$ , the correlation value increases, as the first period lines up with the following period.

Thus, the peaks of  $ACF(w[n])$  can be used to estimate the periodicity of a signal.<sup>1</sup>

In the next step, the peaks in  $ACF(w[n])$  were found using a peak-finding algorithm. Since the analysis window length was more than twice the mean breath length,  $w[n]$  is likely to contain at least two breath cycles, and thus multiple peaks.



**Figure 3** Reference breath length histogram with model normal distribution.

To address the possibility of false alarm peaks produced by this approach, the algorithm models the breath length probabilities with a normal distribution. The parameters  $\mu$  and  $\sigma$  for the normal distribution were calculated using manual breath annotations. The modelled probability distribution is shown in Figure 3 along with a density histogram of the breath lengths from the sets of 8782 manually annotated breaths. Using this probability distribution, the algorithm can rank the breath length candidates, ensuring that it considers the most probable breath length first, thus saving on computing time. Additionally, any breath length candidate whose length probability is less than three standard deviations from the mean is discarded as being too unlikely. Practically speaking, this means that any breath shorter than approximately 1 second or longer than 6 seconds was discarded.

In the next step, for every remaining breath length candidate, a discrete sine template waveform is generated, using the following formula:

$$\sin[n] = \sin\left(\frac{n \cdot 2 \cdot \pi}{l} + \theta\right)$$

Where  $l$  is the length of the candidate in seconds and  $\theta$  is an offset that can be set to  $1.5 \pi$  to shift the waveform so that it starts at  $-1$ , ends at  $-1$ , and has a peak in the middle. To find where a given template waveform fits most closely to the RIP signal window, the algorithm compares it to the RIP signal using the Pearson correlation coefficient ( $\rho$ ) at each point on the RIP signal. The formula for calculating  $\rho$  for a pair of signals is:

$$\rho(w[n], \sin[l]) = \frac{\text{cov}(w[n], \sin[l])}{\sigma_{w[n]} \sigma_{\sin[l]}}$$

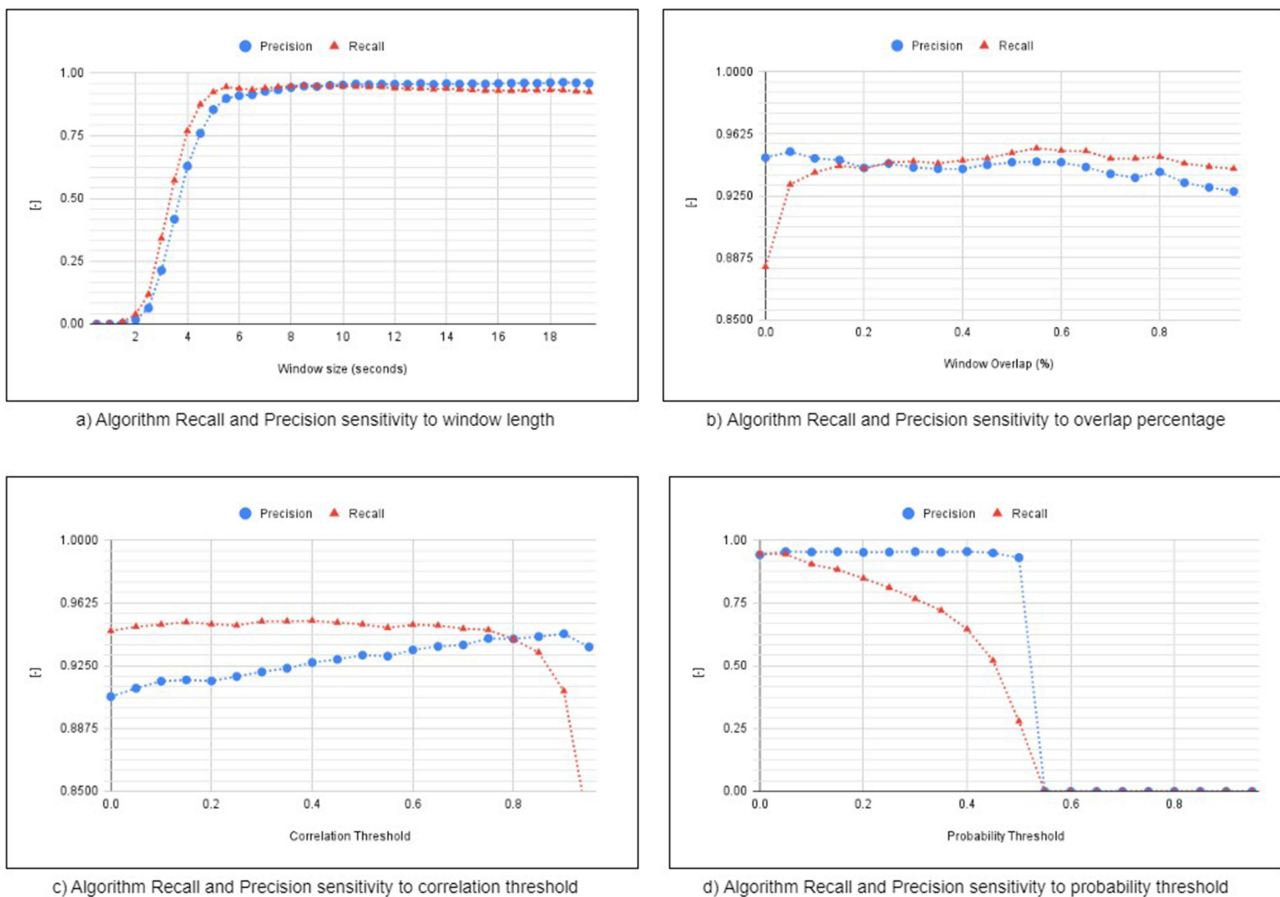
Where  $\text{cov}(w[n], \sin[l])$  is the covariance of the window and template waveform, which can be calculated as:

$$\text{cov}(w[n], \sin[l]) = \frac{\sum_{i=1}^N (w[n][i] - \overline{w[n]}) (\sin[l][i] - \overline{\sin[l]})}{N}$$

where  $N$  is the length of  $w[n]$  and  $\sin[l]$  which must be equal, and  $\overline{w[n]}$  and  $\overline{\sin[l]}$  are the means of the respective signals. The sign of  $\rho$  describes whether the signals are positively or negatively correlated, and its value describes how strong the correlation is. A  $\rho$  value of 0.0 means that the signals are not correlated, a value of 1.0 indicates a positive correlation and a value of  $-1.0$  means that the variables are perfectly inversely correlated. The algorithm treats the RIP signal as one variable and the template waveform signal as another and calculates the correlation of the template waveform over the entire window. The correlation of the template waveform and the RIP signal produces a third signal, whose peaks represent possible onsets of the target breath. Since the template waveform approximates the shape of a breath,  $\rho$  is not expected to reach 1.0. However, the correlation still provides information about the validity of the breath onset. The algorithm discards any breath onset candidate whose  $\rho$  is less than 0.75. The  $\rho$  elimination criterion was chosen via experimentation to eliminate as many inaccurate guesses as possible, while still not being so strict as to eliminate legitimate guesses on noisy data, at approximately 0.5  $\rho$  below the stable elimination criterion (see Figure 4). If this elimination step filters out all breath onset candidates, the algorithm repeats the template waveform fitting process with another breath length candidate. If the algorithm processes all breath length candidates and no breath is found in the current signal window, then the algorithm moves on to the next window. If the correlation is above the threshold, the algorithm adds the onset and the duration to a list of breaths and moves the window onset to the end of the detected breath. This process is repeated until the signal is fully analyzed.

### Breath Placement Post-Processing

As the sliding windows overlap, the algorithm has a tendency to re-discover breaths. To solve this problem, the  $i^{th}$  breath is compared to the  $i+1^{th}$  breath. If the overlap of the breaths spans the majority of the total length, the breaths are



**Figure 4** Algorithm Recall and Precision sensitivity analysis. (a) Window length (b) Overlap percentage (c) Correlation threshold (d) Probability threshold.

considered a double detection, and therefore the detections are merged. The process of merging two breath detections involves replacing them with a single detection which covers the area that both previous detections covered.

The percentage overlap calculation for a pair of time spans is:

$$O_w(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

where  $|A|$  and  $|B|$  are the lengths of time spans A and B respectively, and  $|A \cap B|$  is the overlap area of detections A and B. If the breaths do not overlap at all, the value produced by this function is negative, and in the case of perfect overlap, the overlap value is 1.0. For this reason, the function is clamped above 0.0. The detection merging procedure is visualized in Figure 5a.

Due to the 80% overlap required to merge breaths, the breaths can still overlap by up to 20%.

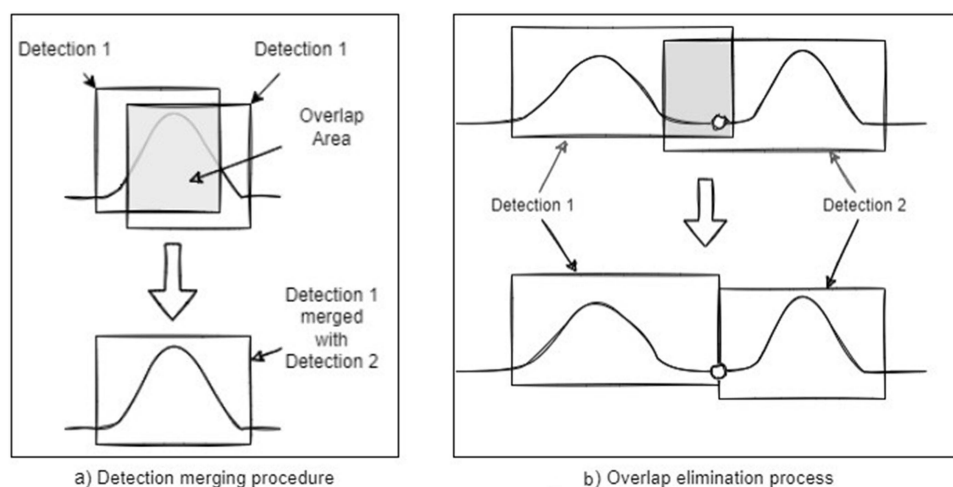
By definition, a breath cannot overlap with another breath, so for each breath, the  $i^{\text{th}}$  breath is compared to the  $i+1^{\text{th}}$  breath. If they still overlap, the end of the  $i^{\text{th}}$  and start of the  $i+1^{\text{th}}$  breath are moved to the time of the minimum value of the RIP signal within the overlapping region. This process is visualized in Figure 5b. The result of this post-processing process is that there is no overlapping pair of detections, satisfying the constraint that no two breaths can share. When run on Evaluation-Subset-A and Evaluation-Subset-B, the post processing step removed 2.05% of detections on average from each interval.

## Algorithm Evaluation

As the algorithm's task is to mark an individual detection event anywhere on a signal, an obvious problem presents itself when comparing detections to a ground truth. If the algorithm produces a false positive, splits a single breath into two or more breaths, or any case in which an extra detection is inserted, a misalignment between the list of detections and annotations is created where the detections placed after the false positive have an index that corresponds to the index of a later annotation than it should. The error compounds after each false positive.

The same misalignment error is created when the algorithm produces a false negative, except the misalignment is now reversed, ie, each detection after the false negative has an index corresponding to the index of an earlier annotation than it should. As with the previous case, the misalignment error compounds after each false negative.

Due to the possibility of misalignment errors, it is impossible to naively compare the list of detections and annotations, and an extra step must be performed to match detections to their corresponding annotations. We solved this alignment problem by using a matrix containing the percentage overlap of all available pairs of detections and annotations calculated using the overlap equation introduced earlier.



**Figure 5** Post processing visualization. (a) Detection merging procedure for double-detections. (b) Overlap removal.

This matrix is referred to as the overlap matrix and simplifies the process of finding which detection corresponds to which annotation, whether a given detection is a false positive or not, and whether a given annotation corresponds to a detection or is a false negative.

Given an overlap matrix  $A$  of a list of detections  $X$  and a list of annotations  $Y$ , the overlap of any  $X[i]$  and  $Y[j]$  can be accessed in  $A[i,j]$ . Using an overlap matrix, a detection corresponding to any annotation could be found by locating the index of the maximum overlap value in the overlap matrix column for that annotation.

To be counted as a correct detection for a given annotation, a detection must have a weighted overlap value of over 80% with that annotation. If an annotation had no value above that threshold in its column in the overlap matrix, the breath was counted as having been missed by the algorithm (false negative). Similarly, if a detection had no value above the threshold in its row in the overlap matrix, the detection was counted as a false positive. Due to the restriction that detections may not overlap, it was impossible for two detections to correspond to the same annotation.

This paper uses the precision (ratio of true positives to true positives and false positives), recall (ratio of true positives to true positives and false negatives), and F1 score metrics to estimate the accuracy of the algorithm. In addition to the precision, recall, and F1, additional statistics were collected on the placement error of the detections that were counted as correct. Those include the length of detections versus the length of the annotations.

The start and end error was calculated using the following two formulae:

$$\text{Start error} = s - s^p \quad \text{End error} = e - e^p$$

where  $s$  is the annotation start,  $e$  is the annotated breath end,  $s^p$  is the predicted breath start, and  $e^p$  is the predicted breath end.

The algorithm has four main parameters; the analysis window length, the overlap threshold, the correlation cut-off for the sine fitting procedure, and the probability threshold for the filtering process. To gauge the effect these variables have on the algorithm's performance, the evaluation was repeated for a range of values for each variable.

## Data Description

An extensive evaluation of the correctness of the algorithms output is required, both during normal breathing, and other conditions that may arise during sleep. The dataset used for validation contained 31 overnight PSGs from people diagnosed with obstructive sleep apnea and people with no known sleep issues (VSN-14-080). Of the participants, 13 were female and 18 were male. The mean age of the participants was 47.1 years, in the range of 20–69 years. The mean body-mass index (BMI) was 29.9 kg/m<sup>2</sup> in the range of 21.6–49.3 kg/m<sup>2</sup>. The mean apnea-hypopnea index (AHI) was 9.3 h<sup>-1</sup> in the range of 0.0 to 34.8 h<sup>-1</sup>. Due either to signal failure in the RIP signal or errors in exporting the recordings from the proprietary NOX format to the standard European data format (EDF), five recordings had to be discarded. Each PSG included all standard signals, including EEG, EOG, EMG, ECG, and airflow recorded with a nasal cannula, thorax and abdomen RIP belts, pulse oximetry (SpO<sub>2</sub>), and an audio signal. The RIP belts in the dataset were recorded with a 25Hz sampling frequency. Additionally, esophageal pressure was recorded with a nose-fed catheter.<sup>22</sup> The algorithm was evaluated against 39 variable-length manually annotated evaluation intervals, which were further split into two evaluation subsets. The first set, referred to as Evaluation-Subset-A, was selected to specifically contain various sleep-disordered breathing (SDB) events, as well as different sleep stages. Evaluation-Subset-A contained 14 variable length intervals with a mean length of 16 minutes, in the range of 1.5 to 37.5 minutes, with a cumulative length of 225.65 minutes (3.6 hours). The SDB events in Evaluation-Subset-A included obstructive apneas, hypopneas, and increases in respiratory effort without apnea or hypopnea. Further events included in Evaluation-Subset-A were sleep stages, movements, oxygen desaturations, and snoring.

These intervals, however, were only selected from one participant in the dataset and are not representative of the general public.

This issue was addressed with a second evaluation subset, referred to as Evaluation-Subset-B, consisting of a collection of 10-minute intervals from the remaining 25 valid PSGs in the dataset, of which 12 participants were healthy (AHI < 5) and 13 had some severity of SDB (AHI mean was 16.4, std. was 8).

These intervals were selected randomly from each recording to avoid cherry-picking favorable intervals. The random selection was done blindly, aside from being restricted from one hour after the recording starts to one hour before the recording ends. This was done to reduce the probability of including either the participants settling down to sleep or moving around as they wake up. The intervals were relatively artefact free, with approximately 90% of the signals in the period being free of artefacts. Due to the requirement that the algorithm be evaluated for its robustness to them, the artefacts were not removed.

The locations of individual breaths in both evaluation subsets were then manually marked using a custom-made scoring tool programmed in Python. The manual breath annotations represented the ground truth.

Of the total 39 intervals in Evaluation-Subset-A and Evaluation-Subset-B, one was found to be incorrectly manually annotated and was discarded. The algorithm was therefore evaluated on 7.3 hours of manually annotated data over 38 intervals, containing 8782 individual breaths from 26 participants.

## Results

The algorithm was evaluated on two sets of manually annotated intervals, the first set containing a relatively high amount of SDB events, and the second set being sampled from a population of 25 participants. Figure 6 shows the format of how the algorithm detects a breath. The performance evaluation results are summarized in Table 1, and the placement errors are shown in Table 2. The algorithm achieved, on average, 0.94 precision for Evaluation-Subset-A and 0.93 for Evaluation-Subset-B. This means that only 6% and 7% of detections were classified as false positives for Evaluation-Subset-A and Evaluation-Subset-B, respectively. The recall for Evaluation-Subset-A and Evaluation-Subset-B was 0.94 and 0.95, respectively, meaning that the algorithm only missed 6% of breaths in Evaluation-Subset-A and 5% of breaths in Evaluation-Subset-B. Two intervals in Evaluation-Subset-A had noticeably worse results, with F1 scores of 0.79 and 0.81. Upon visual inspection, the errors were mainly due to incorrect manual annotations and noise in the signal during those intervals. Omitting these two intervals increased the mean F1 of the algorithm to 0.95 for Evaluation-Subset-B. The algorithm performed noticeably worse for one interval in Evaluation-Subset-A than for the others, its precision being 0.76 and the recall being 0.963, making for an F1 score of 0.854. This lack of performance was due to the interval being relatively short and the beginning of the signal being dominated by a movement event, causing the algorithm to misclassify the movement as breaths.

On the other hand, the algorithm achieved perfect precision for two intervals in Evaluation-Subset-A and one in Evaluation-Subset-B, all of which contained no SDB events and only stable breathing.

The recall was slightly more stable than the precision for both sets, with the standard deviation being 0.054 for the recall and 0.058 for the precision. The mean start error for both Evaluation-Subset-A and B was approximately 6.4% of the mean breath length. The mean end error for both sets was more significant, 10% and 8.4% of the mean breath length for Evaluation-Subset-A and B respectively. When visually inspected, the alignment of the detections and the thoracic RIP signal was high for both Evaluation-Subset-A and B.

## Sensitivity Analysis Results

The analysis window length is the window length in seconds that the algorithm uses to search for breaths at each step. The results of the sensitivity estimation can be seen in Figure 4a and show that both precision and recall rise sharply as the window length reaches approximately 6 seconds and plateaus at approximately 8 seconds. The reason for the sharp rise in performance

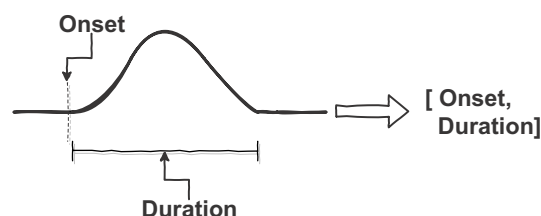


Figure 6 Visualization of an example detection.

**Table 1** Evaluation Results of the RCI Algorithm

Evaluation Subset	Mean Precision	Mean Recall	Mean F1
A	0.94	0.94	0.94
B	0.93	0.95	0.94

**Table 2** Placement Errors of the RCI Algorithm

Evaluation Subset	Annotation Mean Breath Length (Seconds)	Detection Mean Breath Length (Seconds)	Mean Absolute Start Error (Seconds)	Mean Absolute Start Error (Seconds)
A	2.57	2.76	0.16	0.24
B	3.56	3.88	0.23	0.30

between 2–6 seconds is most likely that the window cannot reliably fit two cycles of the respiratory cycle until the window becomes longer than twice the average length of breath in the dataset. The overlap percentage is the amount of the previous window included in the next window as the analysis window advances. The effect of this parameter is shown in [Figure 4b](#), which suggests that the algorithm performs noticeably poorly only in terms of recall when the overlap percentage is around 0. This can be explained by the algorithm missing breaths as the window skips entirely or partly over them. The precision seems largely unaffected, which indicates that the number of false positives drops proportionally with the number of true positives. The best performance in terms of accuracy and recall was around the 55% overlap, which was thus chosen as the default overlap value. The correlation threshold dictates how much a breath candidate must resemble a model breath and is measured by its Pearson correlation (see [Figure 4c](#)). It filters out waveforms that may only superficially resemble breaths but still form peaks in the sine-correlation function. As the correlation threshold increases, the precision improves. This can be interpreted as the criterion for what “looks like a breath” becoming stricter, thus eliminating more false negatives. The recall seems unaffected by this criterion until the threshold reaches approximately 0.8, at which point it sharply drops. This drop in performance is to be expected since the template waveform is only an estimation of the general shape of a breath in the signal, and thus, the correlation with the signal is not expected to be perfect.

The probability threshold parameter is used to discard breaths that are considered too improbable. As [Figure 4d](#) indicates, the absence of this filtering step has little effect. The precision is least affected by the probability threshold, while the recall drops sharply as the threshold increases. This is reasonable since as the threshold increases, more legitimate breaths are discarded, thus negatively affecting the recall until the threshold reaches approximately 0.55, at which point all breaths are discarded. The stability of the precision suggests that the rate of false positives drops proportionally to the rate of true positives as this parameter approaches 0.5. The reason for the falloff of both the precision and recall at 0.5 is that the maximum possible value of the probability estimator is 0.5, so any value above 0.5 will cause the filtering process to discard all detections.

## Discussion

This paper presents a novel algorithm designed to perform RCI on the thoracic RIP signal, based on signal processing and statistical methods. The algorithm achieved an F1 score of 0.94 when detecting breaths during sleep over multiple nights and including SDB events, that is 94% of breaths were classified correctly, with 6% false negatives. Of the detections made by the algorithm, approximately 95% are correctly placed breaths, with only 5% being false positives. This accuracy is superior,<sup>13</sup> or comparable<sup>10,11</sup> with previous work, however, we note that comparison to some prior work is problematic since there is no standardized method of evaluating RCI algorithms, and thus different works approach the task of evaluation differently, making comparisons difficult, if not at times impossible.<sup>12,14,18</sup> Comparison between algorithms can be seen in [Table 3](#). Currently, the algorithm is only evaluated on RIP signals collected with

**Table 3** Comparison Between This Work and Related Work

Algorithm	Accuracy	Signal	N
BreathFinder (This work)	94%	Thorax RIP	8782 respiratory cycles
Rosenwein et al <sup>13</sup>	87%	Audio	188.850 events (inspirations)
Yaha & Faezipour <sup>15</sup>	95%	Audio	128 events
Palaniappan et al <sup>18</sup>	“Very Good”	Audio	72000 events
Hsiao et al <sup>8</sup>	91%	Audio	489 recordings of lung sounds from 22 participants (number of events not specified)
Hult et al <sup>9</sup>	99%-90%	Audio	1863 respiratory cycles
Alshaer et al <sup>23</sup>	Not stated	Audio	–
Moyles & Erlandsson <sup>12</sup>	No Validation	Air Flow	–
Lopez-Meyer et al <sup>11</sup>	96%	RIP Belts	At least 10 minutes from 4 participants (number of events not specified)
ResplnPeace <sup>14</sup>	No validation	RIP Belts	–

**Abbreviations:** PSG, Polysomnography; RIP, Respiratory Inductance Plethysmography; SpO<sub>2</sub>, Oxygen Saturation; EEG, Electroencephalography; EMG, Electromyograms; ECG Electrocardiography; RCI, Respiratory Cycle Isolation; ACF, Auto Correlation function; BMI, Body Mass Index; AHI, Apnea-Hypopnea Index; EDF, European Data Format; SDB, Sleep Disordered Breathing.

a 25Hz sampling frequency. The algorithm is designed to be independent of the sampling frequency of the signal but requires a similarly rigorous evaluation at other sampling frequencies. In the validation data used in this paper, the algorithm is validated on intervals containing significant movement, respiratory events, and various sleep stages.

The evaluation found that the detection rate was not meaningfully influenced by respiratory events, arousals, or physiology, however, the most impactful factor in terms of detection rate seemed to be artefacts caused by movement or signal failure. On the other hand, such artefacts only cause the algorithm to produce errors where the artefacts occur, and cause no errors for future detections, indicating that the algorithm can easily recover from an artefactual period. When the detection error of the correctly detected breaths is expressed as the mean absolute start and end error, the algorithm tends to produce greater end errors than start errors. The mean end error, however, was less than 8% of a mean breath length, and upon visual inspection, was not discernible to the human eye. The start and end errors of both sets may be partially explained by the fact that the manual annotations did not observe the restriction that only one breath can take place at any moment imposed by this work’s definition of the respiratory cycle, effectively introducing small sections of the annotations that at most one detection can overlap with thus artificially negatively impacting the metrics. Although the algorithm was originally designed for use in sleeping individuals, we believe it could be used to research respiration during speech, exercise, emotional response analysis, and other applications provided that the proper evaluation of the output correctness is performed. The total number of participants used for the evaluation of the algorithm was 25. This is comparable to other literature, where the range of the number of individuals used for testing similar tasks ranges between none reported, 4, 75, and 140.<sup>11–14,18</sup> In future work, the algorithm should be evaluated on a much larger dataset. The algorithm achieved a higher accuracy than the AUDAS algorithm,<sup>24</sup> however, the AUDAS algorithm detects individual respiratory phases whereas BreathFinder locates individual respiratory cycles and therefore direct comparison is not appropriate. Similarly to AUDAS, the work done by Hsiao et al achieves 92% accuracy when detecting inspirations and expirations, but as the task is fundamentally different to this approach, direct comparison is not appropriate.<sup>8</sup> The algorithm is designed to work on the thoracic RIP signal, but in theory should also work on the abdominal RIP signal. However, this requires validation to assess the validity of the results. Due to the high detection rate of the algorithm and the relatively low rate of false positives, the authors suggest that the proposed algorithm can be reliably used for future research into the nature of respiration during sleep based on RCI-based adaptive segmentation.

## Clinical Implementation and Applications

The implementation of the BreathFinder algorithm has previously demonstrated its utility in other works, particularly in the identification of Obstructive Apneas. It has been successfully applied in detecting obstructive apneas by using the BreathFinder algorithm on the thoracic RIP signal to find individual breaths, and then performing machine learning on the thoracic, abdominal, and flow signals during those individual breaths, exhibiting impressive performance with a substantial F1 score of 0.94 in apnea detection tasks, thus corroborating its efficacy and reliability in this context.<sup>25</sup> This makes it an instrumental tool in the diagnosis and management of sleep-related disorders. Moreover, the BreathFinder algorithm has shown versatility by its effective application in an unsupervised machine-learning context. Encoding the thoracic and abdominal RIP signals, along with the airflow signal from individual breaths facilitated the exploration of the latent feature space, which consequently allowed the identification of significant clusters of breaths. These clusters demonstrated notable common characteristics, including the incidence of obstructive apneas.<sup>25</sup> This exemplifies the algorithm's capacity for contributing to advanced analytical strategies that expose the intricacies of respiratory patterns. It emphasizes the potential for further exploitation of the BreathFinder algorithm in a myriad of applications, including advanced diagnostics, predictive modeling, and personalized therapeutic approaches.

## Study Limitations

The evaluation has the drawbacks that it is only formally done on the thoracic RIP signal, and the evaluation was only done on data from one dataset. The algorithm has furthermore only been evaluated against a RIP signal using 25 as the sampling frequency. Further research is additionally required to specifically evaluate the effects of events such as changes to body posture, RIP artefacts, incorrect RIP placement, RIP belt stability or other deformations in the signal. The purpose of the algorithm is to isolate breaths, rather than to provide any information or statistics on the nature of the breath further than its location in the signal, and any analysis such as obstructive apnea detection or flow measurement is future work made available by this work.

Due to the low number of breaths that the BreathFinder algorithm is evaluated on, the statistical significance of the results cannot be assured and thus an assessment of the algorithm on larger datasets is needed to evaluate the algorithm's performance when faced with a larger and more diverse range of respiratory events such as central apneas, and therefore, this work can be viewed as a proof-of-concept study.

## Conclusion

This paper introduces BreathFinder, a novel algorithm designed to find individual breaths in the thoracic RIP signal. The algorithm uses periodicity estimation and sine fitting procedures to pinpoint the locations of individual breaths within a PSG.

The algorithm was evaluated on approximately 7.8 hours of manually annotated breathing intervals. The results suggest that the algorithm detects, on average, 94% of breaths correctly, and of the detected breaths, only 4% on average are false positives. The placement error of the correctly detected breaths was generally within acceptable margins, being less than 10% of the mean breath length. The exceptional performance of the algorithm in terms of the evaluation metrics suggests that it is usable for further analysis of sleep data on a breath-by-breath basis. Unlike previous thorax RIP RCI algorithms, BreathFinder is provided as an open-source algorithm, is also validated against a large range of respiratory events, and demonstrates robustness against signal artifacts, also making it the only RCI algorithm evaluated on sleep data known to the authors.

## Public Availability

The algorithm described in this work has been implemented in Python and made open source under the GNU license. The source code is available via GitHub: <https://github.com/benedikthth/BreathFinder>. This paper has been uploaded to Arxiv as a preprint: <https://arxiv.org/abs/2203.01828>.

The library can be installed via the Python package manager: <https://pypi.org/project/BreathFinder>.

## Acknowledgments

We acknowledge the invaluable support and help from Nox Medical.

## Ethics Statement

The National Bioethics Committee (Application: 14-080) and the Data Protection Agency of Iceland approved the study protocol and written consent was obtained from all research subjects. The study adheres to the declaration of Helsinki.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The authors of this paper have received funding to perform this work and write this paper from the European Union's Horizon 2020 research and innovation programme (grant agreement 965417) as well as NordForsk (NordSleep project 90458) via Business Finland (5133/31/2018), the Icelandic Research Fund (ESA & ASI). The Sleep Revolution project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 965417.

## Disclosure

Dr Erna Arnardóttir reports personal fees from Nox Medical, Philips, ResMed, Jazz Pharmaceuticals, Apnimed, Linde Healthcare, and Vistor, outside the submitted work. The authors report no other conflicts of interest in this work.

## References

1. Vlachos M, Yu P, Castelli V. On periodicity detection and structural periodic similarity. In Proceedings of the 2005 SIAM international conference on data mining; 2005:449–460. doi:10.1137/1.9781611972757.40.
2. American Academy of Sleep Medicine. *International Classification of Sleep Disorders*. third ed. 2014.
3. Praetorius HM, Bodenstern G, Creutzfeldt OD. Adaptive segmentation of EEG records: a new approach to automatic EEG analysis. *Electroencephalogr Clin Neurophysiol*. 1977;42(1):84–94. doi:10.1016/0013-4694(77)90153-5
4. Schulz H. Rethinking sleep analysis. *J Clin Sleep Med*. 2008;4(02):99–103. doi:10.5664/jcsm.27124
5. Koch H, Jennum P, Christensen JAE. Automatic sleep classification using adaptive segmentation reveals an increased number of rapid eye movement sleep transitions. *J Sleep Res*. 2019;28(2):e12780. doi:10.1111/jsr.12780
6. Procházková A, Kuchynka J, Yadollahi M, Araujo CPS, Vyšata O. Adaptive segmentation of multimodal polysomnography data for sleep stages detection. In 2017 22nd International Conference on Digital Signal Processing (DSP); 2017:1–4.
7. Chervin RD, Burns JW, Subotic NS, Roussi C, Thelen B, Ruzicka DL. Method for detection of respiratory cycle-related EEG changes in sleep-disordered breathing. *Sleep*. 2004;27(1):110–115. doi:10.1093/sleep/27.1.110
8. Hsiao C-H, Lin T-W, Lin C-W, et al. Breathing sound segmentation and detection using transfer learning techniques on an Attention-Based Encoder-Decoder architecture. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); 2020:754–759.
9. Hult P, Fjällbrant T, Wranne B, Engdahl O, Ask P. An improved bioacoustic method for monitoring of respiration. *THC*. 2004;12(4):323–332. doi:10.3233/THC-2004-12404
10. Hult P, Wranne B, Ask P. A bioacoustic method for timing of the different phases of the breathing cycle and monitoring of breathing frequency. *Med Eng Phys*. 2000;22(6):425–433. doi:10.1016/S1350-4533(00)00050-3
11. Lopez-Meyer P, Sazonov E. Automatic breathing segmentation from wearable respiration sensors. In 2011 Fifth International Conference on Sensing Technology; 2011:156–160. doi:10.1109/ICST.2011.6136953.
12. Moyles TP, Erlandson RF, Roth T. A nonparametric statistical approach to breath segmentation. In Images of the Twenty-First Century. Proceedings of the Annual International Engineering in Medicine and Biology Society; 1989. vol.1:330–331. doi:10.1109/IEMBS.1989.95756.
13. Rosenwein T, Dafna E, Tarasiuk A, Zigel Y. Detection of breathing sounds during sleep using non-contact audio recordings. *Conf. Proc. IEEE Eng Med Biol Soc*; 2014:1489–1492.
14. Włodarczak M. RespInPeace: toolkit for processing respiratory belt data; 2019. doi:10.5281/ZENODO.3246019.
15. Yahya O, Faezipour M. Automatic detection and classification of acoustic breathing cycles. In Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education; 2014:1–5. doi:10.1109/ASEEZone1.2014.6820648.
16. Korten J, Haddad G. Respiratory waveform pattern recognition using digital techniques. *Comput Biol Med*. 1989;19(4):207–217. doi:10.1016/0010-4825(89)90009-7
17. Thordarson B, Islind AS, Arnardóttir E, Óskarsdóttir M. Exploration of sleep events in the latent space of variational autoencoders on a Breath-by-Breath basis. In Proceedings of the 56th Hawaii International Conference on System Sciences (Hawaii: Computer Society Press; 2023:3091–30911.
18. Palaniappan R, Sundaraj K, Sundaraj S. Adaptive neuro-fuzzy inference system for breath phase detection and breath cycle segmentation. *Comput Methods Programs Biomed*. 2017;145:67–72. doi:10.1016/j.cmpb.2017.04.013
19. Portier F, Portmann A, Czernichow P, et al. Evaluation of home versus laboratory polysomnography in the diagnosis of sleep apnea syndrome. *Am J Respir Crit Care Med*. 2000b;162(3):814–818. doi:10.1164/ajrccm.162.3.9908002
20. BaHammam AS. Signal failure of type 2 comprehensive unattended sleep studies in patients with suspected respiratory sleep disordered breathing. *Sleep Breath*. 2005;9(1):7–11. doi:10.1007/s11325-005-0001-6

21. Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. *Analy Chem.* 1964;36(8):1627–1639. doi:10.1021/ac60214a047
22. Serwatko M. Validation of a new method to assess respiratory effort non-invasively [Master's thesis], Reykjavik University; 2016.
23. Alshaer H, Fernie GR, Sejdić E, Bradley TD. Adaptive segmentation and normalization of breathing acoustic data of subjects with obstructive sleep apnea. In 2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH); 2009:279–284. doi:10.1109/TIC-STH.2009.5444489.
24. Lalouani W, Younis M, Emokpae RN, Emokpae LE. Enabling effective breathing sound analysis for automated diagnosis of lung diseases. *Smart Health.* 2022;26:100329. doi:10.1016/j.smhl.2022.100329
25. Þórðarson BH. Analysis and detection of obstructive apnea in individual breath cycles [Master's thesis]. Reykjavik University; 2021.

Nature and Science of Sleep

Dovepress

**Publish your work in this journal**

Nature and Science of Sleep is an international, peer-reviewed, open access journal covering all aspects of sleep science and sleep medicine, including the neurophysiology and functions of sleep, the genetics of sleep, sleep and society, biological rhythms, dreaming, sleep disorders and therapy, and strategies to optimize healthy sleep. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/nature-and-science-of-sleep-journal>