Accuracy-Coverage Tradeoff of Nocturnal Vital Sign Estimation in Smart Beds

Daniel Waltisberg
Wearable Computing Lab
ETH Zürich
wdaniel@ife.ee.ethz.ch

Oliver Amft
ACTLab
Chair of Sensor Technology
University of Passau
amft@ieee.org

Gerhard Tröster
Wearable Computing Lab
ETH Zürich
troester@ife.ee.ethz.ch

Abstract
We introduce a novel evaluation approach for smart bed systems that continuously measure vital signs. In particular, we demonstrate that estimation accuracy (or error) and measurement coverage time are key performance metrics, describing a performance tradeoff in practical smart bed systems. Based on a typical smart bed system that uses a force sensor array placed between bed mattress and frame, we evaluate the effect of different signal filtering options to illustrate viable design choices using our accuracy-coverage tradeoff analysis. In a full night recording study with six participants focusing on respiration rate estimation, we show that measurement coverage is an essential metric that should be analysed together with accuracy, when assessing the performance of smart bed systems.

Author Keywords
Sleep; Respiration rate; Accuracy; Coverage

ACM Classification Keywords
J.3 [Life and Medical Sciences]: Consumer Health

Introduction
Vital signs monitoring during the night is essential for health assessment and patient monitoring. For example, respiration rate is estimated using nasal monitors and face
spirometers in clinical sleep analysis laboratories, among other equipment, to assess sleep abnormalities such as apneas. Furthermore, continuous respiration rate can be used to estimate sleep phases [4]. Adequate proportions of the different sleep phases are frequently considered as sleep quality measure [2]. Since clinical sleep laboratories are only available for limited monitoring periods and uncomfortable due to the numerous measurement devices, a trend to ubiquitous smart bed systems has recently emerged. Various approaches based on force, piezoelectric, and pneumatic sensors have been proposed for sleep and respiration monitoring, e.g. [13, 14, 11, 1]. While several approaches were made to estimate vital parameters from bed-integrated sensors, full nocturnal measurements in different beds have not been sufficiently investigated (see related work for more details). In particular, natural body movement during sleep can critically affect the system accuracy. In contrast, when considering short time segments of several minutes only, near to perfect accuracy could be achieved since test users lie still on the bed. Therefore we consider the tradeoff between vital sign estimation accuracy and measurement coverage during sleep at night as an open issue. In particular, since many smart bed systems use sensor arrays and different sensor signal processing methods, the effects of signal filtering, fusion methods, and parameter settings on the accuracy-coverage tradeoff should be further investigated.

In this paper, we present an accuracy-coverage tradeoff analysis approach to investigate the overall viability of a typical smart bed system using a portable force sensor array place below the bed mattress. Our work focuses on providing realistic performance indicators. Using the accuracy-coverage analysis, we investigate the effect of various signal filtering and fusion design options. In particular, the paper provides the following contributions:

1. We introduce measurement coverage as additional performance metric besides the typically considered accuracy or error metrics. We argue in this paper that the metrics pair (accuracy/error and coverage) needs to be jointly analysed to determine the overall smart bed performance. We show that the performance tradeoff can be nicely illustrated using root-mean-square-error (RMSE) vs. coverage plots.

2. We use our accuracy-coverage analysis to assess respiration rate estimations in a night sleep study with six participants using a bed-integrated force sensor unit. A wearable reference system was used to assess performance.

3. Finally, we also show that our smart bed system can be used to indicate sleep phases, suggesting that the applied signal processing procedures are suitable for the application in sleep monitoring.

Related Work
Various investigations monitored respiration rate and similar vital signs using smart bed systems, but the evaluation methods and derived information quality vary. The most extensive system evaluations have been done by Mack et al. [6] and Zhu et al. [14]. Some other works presented only short evaluation durations, during which subjects were able to keep from moving [1, 7] or did not include reference measurements [5]. In further evaluations, the signals were annotated manually [11] during short time periods and the feasibility of respiration rate estimation was only shown by an analysis of signal to noise ratio [13, 8].
Coverage metric
Only a few works actually reported measurement coverage, i.e. the time during which a respiration rate estimation is not feasible due to low signal quality, e.g. caused by movement. In Mack et al. [6], 16.5% of the test set had to be removed due to movement artefacts and in Zhu et al. [14] it were about 5%. In Shin et al. [11] removed data was calculated for one subject for seven days and ranged between 5% and 10%. We argue that the respiration rate estimation has not been sufficiently investigated in literature, especially the tradeoff between measurement coverage and estimation accuracy.

Respiration rate estimation
In this paper we focus on the estimation of the respiration rate from respiration signals as an example vital sign parameter that can be acquired through a smart bed system. For the rate estimation several approaches have been proposed in literature. Zhu et al. [14] proposed the detection of upward zero-cross points as characteristic points in the respiration signal and compared the points counted within each minute with the reference from a nasal thermistor. In an experiment with 13 subjects and a total recording time of about 24 hours, they achieved a sensitivity and positive predictability of 95.63% and 95.42%, respectively. However, it should be noted that no rates were estimated during artefacts in the respiration signal, which was in about 5% of the data. Another similar method in [1] uses a moving window and determines the maxima and minima in the signal, but only a small evaluation study was performed.

In Mack et al. [6], the initial estimates from peak detections were further adjusted with an algorithm employed over several episodes. In overnight studies with 40 people the standard deviation of discrepancies, i.e. the RMSE, was 2.10 breaths per minute (bpm) for all participants and 1.59 bpm for participants without sleep apnea. However, about 16.5% of the test set had to be removed due to movement artefacts.

While the zero-crossing and peak detection methods allow to calculate the instantaneous rate, they can easily be affected by small fluctuations and disturbances. In [13] the fast Fourier transform was used to analyse the signals in the frequency spectrum, but the time-frequency uncertainty principle limits the accurate calculation of an instantaneous rate. With a window length of 51.2 s the respiration resolution was 1.17 counts per minute.

For our system we used the autocorrelation method described in [7] and [5] to estimate the respiration rate. With an analysis window length of 15 to 30 s, the autocorrelation method was shown to provide sufficient sensitivity in respiration rate computation to estimate REM sleep [4].

Key evaluation metrics
The key evaluations metrics for our tradeoff analysis are coverage and accuracy. The coverage $c$, which compares the number of reliable periods $M$ to the total number of periods $N$, is defined as:

$$ c = \frac{M}{N}, \quad \text{with} \quad M = \sum_{n=1}^{N} b(n), $$

where $b(n)$ is a binary value that represents the reliability of the rate estimation. Our second key metric is accuracy, which we describe with the root mean squared error:

$$ \text{RMSE} = \sqrt{\frac{1}{M} \sum_{n=1}^{N} b(n) (R_{\text{ref}}(n) - R_{\text{est}}(n))^2}, $$
where \( RR_{est}(n) \) is the estimated respiration rate and \( RR_{ref}(n) \) the reference respiration rate. The RMSE is only calculated for periods which are reliable.

**System overview and implementation**

In this work, we used a portable sensor unit from compliant concept\(^1\). The sensor unit was placed between bed mattress and bed frame, see Figure 1. The sensor unit has been developed for the usage in hospital beds, but it can also be used on a normal bed with a slatted frame. The pressure signals are sampled at a rate of 100 Hz and stored at the bedside monitor.

The processing of the recorded sensor signals was adapted from [5] and is shown in the block diagram in Figure 2. For each sensor \( i \) the signal processing is done individually. After filtering, the respiration rate \( RR \) and the reliability measure \( r \) are estimated with the autocorrelation method. In the reliability evaluation block, the binary reliability value \( b_i \) is estimated based on \( RR_i, r_i, \) the signal variance \( \sigma^2 \) and some threshold settings. The resulting binary reliability value \( b_i \) is 1 for a reliable signal and 0 otherwise.

Finally, the respiration rate estimations from each sensor are combined in the sensor fusion block.

**Filtering**

For an adult in rest, the expected respiration rates are between 12 and 30 breaths per minute (bpm) [9] which corresponds to a frequency between 0.2 Hz and 0.5 Hz. For the extraction of the relevant respiration waveform, different filter frequencies and designs have been used in literature. Since the success of the filtering is highly dependent on the sensor technology and signal quality, we implemented three different filtering methods:

- **Butterworth**: The butterworth filter has already been used in previous systems [11]. We applied a 10th-order Butterworth bandpass filter with cutoff frequencies at 0.1 and 1 Hz.

- **Wavelet**: The wavelet transformation has been previously used for the extraction of the respiration waveform [14, 3, 12]. We applied the wavelet multiresolution decomposition until the 8th level with the Daubechies mother wavelet of level 4. From the detail coefficients \( d_7 \) and \( d_8 \) solely, the inverse steps of the decomposition are done.

- **Notch**: The main difficulty to retrieve the respiration waveform is to remove the DC offset of our signals. We therefore used a FIR lowpass filter with cutoff frequency at 1 Hz to remove the high frequency components in combination to a 0 Hz notch filter with a cutoff frequency at 0.03 Hz.

**Autocorrelation**

The autocorrelation method has been previously used in [5] and [7] for the estimation of the respiration rate. Each respiratory waveform \( y \) was segmented into 15 s windows with 10 s overlap - resulting in an update every 5 s. The window size of 15 s ensures that the signal contains at least two respiration cycles.

The normalized autocorrelation was calculated for each of the segmented waveforms. At a lag of one respiration cycle, between 2 s and 6 s, we expect a peak in autocorrelation coefficients due to the sinusoidal waveform of the respiration. We therefore search for the time lag with maximal autocorrelation coefficient in this time window and calculate the corresponding respiration rate \( RR \) from the inverse of the time lag. To describe the confidence of the calculated respiration rate, we use the

---

\(^1\)http://compliant-concept.ch/en/
value of the maximal autocorrelation coefficient as our reliability measure $r$. The maximal value of our reliability measure is therefore 1, which could be reached for a perfectly sinusoidal waveform.

**Reliability Evaluation**
We assume that the calculated respiration rate $RR$ is a reliable result if a) the reliability measure $r$ is above a defined threshold $\theta_r$ and b) the variance $\sigma^2$ is below a defined threshold $\theta_\sigma$. The reliability threshold $\theta_r$ describes the allowed irregularities of the respiration waveform and the variance threshold $\theta_\sigma$ is used for the detection of segments with high variance which usually corresponds to body movements. During evaluation, a parameter sweep of these two thresholds will be done.

**Sensor Fusion**
The above calculations are performed for each sensor $i$ individually. The sensor results $(RR, r, b)_i$ can be fused to improve the respiration rate accuracy and coverage. If at least one sensor yield a reliable rate, the binary reliability output $b$ was set to 1 and the respiration rates were fused. We used the weighted average of the reliable respiration rate, but any other fusion method could be applied.

**Evaluation Study**
We conducted a sleep study to evaluate our system and show the advantage of the coverage versus accuracy tradeoff. We recruited six PhD students (5m, 1f) from our lab with an average age of 26.8 years (SD: 2.6 years). The subjects were recorded during one to three nights at their homes. All subjects could sleep in their normal environment. We recorded 10 nights with an average recording duration of 6.8 h (SD: 1.5 h) and a total of more than 76 hours. After visual inspection, we had to remove one night from the evaluation set due to a visible respiration waveform of a second person in the double bed.

In order to minimally influence the sleep behaviour, we refrain from complex polysomnographic recordings and used the belt system BioHarness™ 2 from Zephyr® as reference system. In our setting, we recorded the respiration waveform for the calculation of the reference respiration rate and the acceleration signal from the belt for the manual synchronisation with the bed sensor.

To calculate the reference respiration rate, we implemented the detection of upward zero-cross points described in [14] and did visual adjustments. About 11% of the recording had to be removed due to distortion in the reference signal through body movements. The average respiration rate from the manual annotations was computed every 5 s with a window length of 15 s.

**Results**
In Figure 3, a representative example of the estimated respiration rate and the binary reliability output is shown and compared with the reference rate from the Zephyr respiration belt. The reference respiration rate shows high fluctuation which is similarly reflected in the estimated respiration rate.

**Comparison of Filtering Methods**
Figure 4 depicts the coverage versus accuracy tradeoff for different filtering methods. For each filtering method, a parameter sweep of the reliability settings was done and the coverage and root mean squared error (RMSE) were computed for each setting. The best achievable coverage is limited to 0.896 by the reference coverage. The notch filtering shows the best results, followed by the butterworth and wavelet filtering. At an 80% coverage, 2

---

2http://www.zephyranywhere.com/
the expected RMSE are 0.75 (Notch), 0.80 (Butterworth) and 0.84 (Wavelet) breaths per minute.

Figure 4: Coverage versus accuracy tradeoff for different filtering methods. Best achievable coverage is limited to 0.896 by the reference coverage. In this plot, ideal performance corresponds to a coverage of 1 and RMSE of 0.

Fusion Improvement
In Figure 5, the tradeoff is shown for single sensors, the average sensor and the fusion result. The coverage and accuracy values of the average sensor were calculated by taking the mean of all sensor metrics for each threshold setting. The fusion result clearly outperforms the average sensor and each individual sensor for coverage and accuracy. At a coverage of 70%, the fusion estimation has a RMSE of 0.61 bpm, whereas the average sensor has a RMSE of 1.04 bpm.

Figure 5: Evaluation comparison of individual sensor rates, the rate of the average sensor and the fusion rate.

Participants
For each participant, the coverage versus accuracy tradeoff was analysed, see Figure 6. The results show individual differences, but show the general tradeoff of coverage and accuracy. The best results were achieved for Participant 4.

Figure 6: Evaluation comparison for different study participants with the notch filter method for filtering.

REM Sleep Estimation
To demonstrate the high time resolution sensitivity of the rate estimations, we used the variance in the estimated respiration rate for the detection of REM sleep. It is known in literature that the respiration rate has a higher variance during REM sleep [4].

Our reference device for the detection of REM sleep was the ZEO Bedside Monitor. It consists of a headband which measures the EEG with three electrodes. The ZEO does not influence with the respiration reference and the bed sensor unit. The accuracy of the ZEO Bedside monitor has been evaluated in [10] and the REM sleep detection showed an agreement of 85% and 79% with two expert scorers.

We used the standard deviation of the respiration rate with a manually adjusted threshold to estimate REM sleep episodes for one participant. The result in Figure 7 shows a good agreement with the reference from the ZEO.

Figure 7: Representative example of REM sleep stage estimation using respiration rate variation.
Discussion
With a dataset of more than 76 hours of recording time, we could demonstrate advantages in analysing the tradeoff between measurement coverage and RMSE. The tradeoff visualisation allowed us to better compare different algorithm options, in particular by comparing different filtering methods and the effect of the sensor fusion. While in earlier work, measurement coverage was only rudimentary considered, our results clarified that smart bed systems need to be simultaneously assessed by an accuracy and coverage metric to determine suitable system performance.

The advantage of portable or wearable devices as reference measurement is that sleep monitoring can be performed outside of a sleep laboratory, i.e. in the user’s home. For the present study, the wearable system was beneficial as the unnatural sleep lab environment could be avoided. Nevertheless, wearable device could influence sleep behaviour too, e.g. if the device is uncomfortable to wear when lying. Wearable systems for sleep monitoring often provide limited reference signal quality. Hence accuracy, but also measurement coverage affects wearable systems as reference measurement too. In our evaluation, the maximum measurement coverage was limited by the performance of the Zephyr belt.

With the REM sleep estimation results, we could further show that the presented system is sensitive enough to capture small respiration rate fluctuations. Further research should show if the respiration rate could detect REM sleep in a larger population.

The participants in our study were young and healthy. Further research is needed to investigate how the accuracy-coverage tradeoff changes for an older population that potentially exhibits more interrupted sleep. While age is not expected to change respiration rate, movement may well increase during sleep interrupts. Hence the tradeoff analysis has further potential to describe system performance and feasibility for different age groups.

Conclusion and Future Work
In this paper we presented a novel evaluation approach that considers both measurement accuracy and measurement coverage in time for the evaluation of the respiration rate estimation of an unobtrusive bed system. We showed how the accuracy-coverage tradeoff can provide essential insight into the system performance and performance of different algorithm choices. By analysing the metrics pair, it will be feasible to compare the performance of smart bed systems during full night recordings. We expect that the accuracy-coverage analysis can be similarly applied to other vital signs, such as heart rate.

Being an unobtrusive monitoring approach, our smart bed system showed to be practical for full night recordings and could be used during longitudinal studies. As our supplementary comparison of the continuous respiratory rates in comparison to REM sleep phases illustrated, estimating sleep phases is a viable objective for smart bed systems.

Acknowledgements
The authors are grateful for participation of the subjects and for the sensor units which were generously provided by compliant concept.

References
Contact-free measurement of heart rate, respiration


