EINDHOVEN UNIVERSITY OF TECHNOLOGY

Video Respiration Monitoring

Citation for published version (APA):
Lorato, I. R. (2021). *Video Respiration Monitoring: Towards Remote Apnea Detection in the Clinic*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Electrical Engineering]. Eindhoven University of Technology.

Document status and date:
Published: 19/11/2021

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 05. Oct. 2023

Video Respiration Monitoring

Towards Remote Apnea Detection in the Clinic

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een commissie aangewezen door het College voor Promoties in het openbaar te verdedigen op vrijdag 19 november 2021 om 11.00 uur

door

Ilde Rosa Lorato

geboren te Potenza, Italië

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr.ir. M. Matters-Kammerer
promotor:	prof.dr.ir. G. de Haan
1^e copromotor:	dr.ir. S. Stuijk
2^e copromotor:	dr.ir. C. van Pul
leden:	prof.dr.ir. D. Merhof (RWTH Aachen University)
	prof.dr.ir. R.M. Aarts
	prof.dr.ir. P.H.N. de With
adviseur:	dr. P. Andriessen (Máxima Medisch Centrum)

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Video Respiration Monitoring

Towards Remote Apnea Detection in the Clinic

Ilde Rosa Lorato

Doctorate committee:

prof.dr.ir. G. de Haan	Eindhoven University of Technology, promotor
dr.ir. S. Stuijk	Eindhoven University of Technology, 1st copromotor
dr.ir. C. van Pul	Eindhoven University of Technology, 2nd copromotor
prof.dr.ir. M. Matters-Kammerer	Eindhoven University of Technology, chairman
prof.dr.ir. R.M. Aarts	Eindhoven University of Technology
prof.dr.ir. P.H.N. de With	Eindhoven University of Technology
prof.dr.ir. D. Merhof	RWTH Aachen University
dr. P. Andriessen	Máxima Medisch Centrum

This work is part of the Alarm-Limiting AlgoRithm-based Monitoring (ALARM) project funded by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) with grant number 15345. The research took place in the Eindhoven MedTech Innovation Centre (e/MTIC) framework, as collaboration between the Eindhoven University of Technology, Máxima Medical Centre, and Philips Research.

Cover design by Ilde Lorato. The license for the image on the back of the cover was acquired from Adobe Stock, © EkaterinaVladimirova/stock.adobe.com

Printed by Gildeprint, Enschede, The Netherlands

A catalogue record is available from the Eindhoven University of Technology Library
ISBN: 978-90-386-5371-6

© Copyright 2021, Ilde Lorato
All rights reserved. Reproduction in whole or in part is prohibited without the written consent of the copyright owner.

Summary

Respiration rate (RR) and respiration irregularities are often the earliest indicators of physiological deterioration. Furthermore, respiratory disorders can be characterized by cessations of breathing, i. e. apneas. Consequently, in some hospital environments, such as critical care and sleep clinics, respiration is continuously monitored using different sensors and electrodes. In sleep clinics subjects affected by sleep-disordered breathing (SDB), where obstructive sleep apnea (OSA) is one of the most common disorders, are diagnosed using polysomnography (PSG), which monitors several vital signs. Between these, respiratory motion (effort) and flow are monitored using respiratory belts and flow sensors (thermistors and/or pressure transducers) positioned close to the nostrils and/or mouth. The many sensors and electrodes required and their wiring can cause discomfort in patients and interfere with their sleep. Premature babies, instead, are cared for in neonatal intensive care units (NICUs). Apnea of prematurity (AOP) is a common developmental disorder in this population and it is characterized by the presence of apneas. Here, respiration is monitored through chest impedance (CI) that, by using the same electrodes used to measure the electrocardiogram (ECG), monitors changes in the thoracic impedance reflecting the breathing mechanics. The sensitive skin of premature infants can suffer from irritation or even damage due to the use of adhesive electrodes.

Unobtrusive solutions for respiration monitoring are being investigated aiming at reducing the discomfort of patients. Between the unobtrusive alternatives (e. g. radars, pressure-sensitive films, and audio) cameras are an attractive option. They can provide contextual information and images of the patients, which can be useful to the caregivers for patient observation. Camera-based respiration monitoring still faces several challenges as its application in complex hospital environments is not straightforward. In this thesis, we investigate the use of cameras as a substitute for the current respiration monitoring technologies. In particular, we proposed solutions for improved apnea detection and classification, automatic respiratory pixels identification, and non-respiratory motion robustness and analysis. Furthermore, practical challenges have also been considered, such as the possibility of using low-cost cameras to develop a multi-camera setup and the use of cameras in real clinical settings, bringing the technology closer to the clinic.

Firstly, in Chapter 2 the necessity of strategies for the detection of a cessation of breathing (COB), or even an apnea, with low latency usable in camera-based respiration signals, was addressed. The algorithm was developed to detect sudden amplitude changes in the respiration waveform. The proposed method was applied

to the respiration signals extracted from a set of infants' video recordings collected with red-green-blue (RGB) and near-infrared (NIR) cameras. A method from literature was used to extract the respiration signal from the videos. Our method successfully detected short COBs, that were present in the videos, and obtained comparable results to the benchmark method with the difference that our solution can detect COBs with low latency.

Next, in Chapter 3 we analyzed the usability of a thermopile array as a low-cost alternative to thermal cameras for the detection of respiratory flow. Camera solutions in the thermal domain would offer several advantages in clinical settings, e. g. absence of illumination problems and visibility of both respiratory motion and flow. However, their still high cost combined with the clear necessity of multi-camera setup prevents this technology from being considered as a viable option. A feasibility study was conducted to analyze the performance of the thermopile array using videos collected on healthy adults. Different simulated respiratory patterns and rates, as well as spontaneous breathing, were included. The thermopile array performed well in the ideal conditions of the study, but limitations are highlighted due to the extremely low-resolution of the setup (8×8 pixels).

In Chapter 4 a respiration detection algorithm that automatically identifies the pixels containing respiration information from a combination of the three camera views is described. The method was developed and tested using low-resolution thermal videos acquired in a neonatal ward. Moving to a clinical environment, i. e. the neonatal ward, the setup used was adapted to reflect possible complications, we used three low-resolution thermal cameras (60×80 pixels each) to look at the infants from different positions. The use of low-resolution devices allows multi-camera setups avoiding reaching prohibitive costs and prevents possible privacy complications. However, face or body landmark detectability, already complex in hospital environments and infants, becomes further challenging. The algorithm proposed, to automatically identify respiratory pixels, is independent of facial and body landmark detection. The method was benchmarked against a similar method from literature and the superior performance and adaptability of our solution were shown. In this first study, only the moments in which the infants were not moving were used, proving that the method can successfully work in the absence of movement, and highlighting the necessity of motion robustness strategies.

Chapter 5 responds to the clinical expectation of usability and to the limitation of the previous solution by analyzing the RR monitoring algorithm during different motion events in the infants' thermal videos. The method's usability was extended to non-thermal videos as well. Severe movements of the infants were automatically identified from the videos and are not included in the detection of respiration, reflecting the limitation of the current monitoring technology, i. e. CI, which often also fails to give an accurate RR during these moments. Minor movements of the infants were included in the analysis and the performance in the detection of the RR during these events, during the absence of motion, and during non-nutritive-sucking (NNS) was compared. The inclusion of these segments allows to increase

the usability of the method increasing the similarity with the CI and proving the improved motion robustness of our solutions, but limitations on the accuracy of the RR are highlighted.

In Chapter 6 a method to automatically identify the pixels containing specifically respiratory flow in infants' thermal videos without relying on facial landmark detection is proposed. Mixing respiratory motion and respiratory flow is problematic in the case of apneas with an obstructive component, i. e. with respiratory effort present. The algorithm proposed in Chapter 4 and Chapter 5 still presented this drawback. Furthermore, obstructive apneas (OAs) are simulated to compare their detectability between a respiration signal obtained by using only the respiratory flow and one obtained by mixing respiratory flow and motion. The apnea events are detected using the method proposed in Chapter 2. Finally, in Chapter 7 the long-term societal impact of this work is analyzed for the patients and their families, for the clinicians, and for industry. Knowledge utilization is also detailed.

In summary, this thesis analyzes the possibility of using camera-based respiration monitoring in clinical settings aiming at apnea detection. Overall, further research is required to evaluate the accuracy of respiration monitoring using cameras in the clinic. Combining multiple contactless technologies could be the best way to tackle practical problems typical of complex environments, and extend the range of vital signs detectable without contact with the patients.

Acronyms and abbreviations

ACC accuracy

AHI apnea-hypopnea index

AOP apnea of prematurity

AR autoregressive

BalACC balanced accuracy

BPM breaths per minute

BW birth weight

CA central apnea

CI chest impedance

COB cessation of breathing

CSA central sleep apnea

DFT discrete Fourier transform

DOF difference of frames

ECG electrocardiogram

EMG electromyography

FFT fast Fourier transform

GA gestational age

HR heart rate

HSAT home sleep apnea testing

IRT infrared thermography

LWIR long-wave infrared

MA mixed apnea

MAE mean absolute error

MMC Máxima Medical Centre

MR mixed-respiratory

MWIR mid-wave infrared

NaN not a number

NICU neonatal intensive care unit

NIR near-infrared

NNS non-nutritive-sucking

OA obstructive apnea

OF optical flow

OSA obstructive sleep apnea

PA postnatal age

PB periodic breathing

PMA postmenstrual age

PPG photoplethysmography

PR percentage of correct estimates

PSG polysomnography

PT percentage of time used

RE relative error

RefRF reference respiratory flow

RF respiratory flow

RGB red-green-blue

RIP respiratory inductance plethysmography

RM respiratory motion

RMSE root mean square error

ROC receiver operating characteristics

ROI region of interest

RR respiration rate

SD spectral density

SDB sleep-disordered breathing

SE sensitivity

SIDS sudden infant death syndrome

SNR signal to noise ratio

SP specificity

SQI signal quality index

STFT short time Fourier transform

TFN time false negative

TFP time false positive

TTN time true negative

TTP time true positive

VRM video respiration monitoring

Contents

Summary **i**

Acronyms and abbreviations **v**

1 Introduction **1**
1.1 Respiration 1
1.1.1 Respiratory patterns and apnea 2
1.2 Respiration monitoring 4
1.2.1 Traditional methods 4
1.2.2 Unobtrusive alternatives 8
1.3 Problem statement and objectives 10
1.3.1 Problem statement 10
1.3.2 Objectives 11
1.4 Thesis outline 11

2 On-line cessations of breathing detection **13**
2.1 Introduction 14
2.2 Materials and methods 16
2.2.1 Materials 16
2.2.2 Methods 17
2.3 Results 21
2.4 Discussion 22
2.5 Conclusions 24

3 Respiratory flow monitoring using a thermopile array **25**
3.1 Introduction 26
3.2 Materials and methods 27
3.2.1 Materials 28
3.2.2 Methods 30
3.3 Results 34
3.4 Discussion 35
3.5 Conclusions 41

4 Multi-camera infrared thermography and automatic respiratory pixels identification 43
4.1 Introduction 44
4.2 Materials and methods 45
4.2.1 Materials 46
4.2.2 Methods 48
4.3 Results 54
4.4 Discussion 56
4.5 Conclusions 59

5 Camera-based respiration monitoring and motion analysis 61
5.1 Introduction 62
5.2 Materials and methods 63
5.2.1 Materials 63
5.2.2 Methods 66
5.3 Results 75
5.4 Discussion 77
5.5 Conclusions 82

6 Automatic separation of respiratory flow from motion in thermal videos 83
6.1 Introduction 84
6.2 Materials and methods 85
6.2.1 Materials 86
6.2.2 Methods 90
6.3 Results 96
6.4 Discussion 98
6.5 Conclusions 101

7 Societal Impact 103
7.1 Knowledge Utilization 104
7.1.1 Patent application 1 104
7.1.2 Patent application 2 106

8 Conclusions and future research directions 107
8.1 Conclusions 107
8.2 Future research 110

Bibliography 113

Acknowledgements **131**

Curriculum Vitae **133**

List of publications **135**

CHAPTER 1

Introduction

1.1 Respiration

The primary function of the respiratory activity is gas exchange. Oxygen is transferred to the bloodstream and tissues, whereas carbon dioxide is removed from the body. This is possible thanks to the respiratory system, which includes the upper airways (nose, pharynx, and larynx) and the lower airways (trachea, bronchial tree, and lungs) [1].

Diaphragm and external intercostal muscles are the main actors in the mechanics of breathing, through their contraction and relaxation the volume of both thoracic cavity and lungs changes, causing variations in the internal lungs pressure [2]. Two main phases are generally defined, as shown in Figure 1.1, and detailed below.

- Inspiration is initiated by the contraction of the diaphragm and external intercostal muscles, and followed by an increase in the volume of the thoracic cavity and lungs. The pressure in the lungs decreases allowing air to move into the lungs following the pressure gradient [3]. The main effect is inhaling air, ideally through the nasal cavities, where thanks to the underlying blood vessels the inhaled air nearly saturates with water vapor and warms up reaching almost the body temperature.

- Expiration follows the inspiration phase and it is a passive process. It is caused by the relaxation of the inspiratory muscles. This results in a decrease in the thoracic cavity's and lungs' volumes, producing an increase of pressure inside the lungs [3]. Thus, the air is driven out of the lungs. The exhaled air is saturated with water and warm, close to the body temperature, and will lose some of its humidity and warmth to the mucous membrane.

Respiration is automatically controlled by the respiratory control center in the brain [2]. The respiration rate (RR), usually indicated as the number of breaths per minute, is one of the main vital signs. The range of a normal RR at rest is

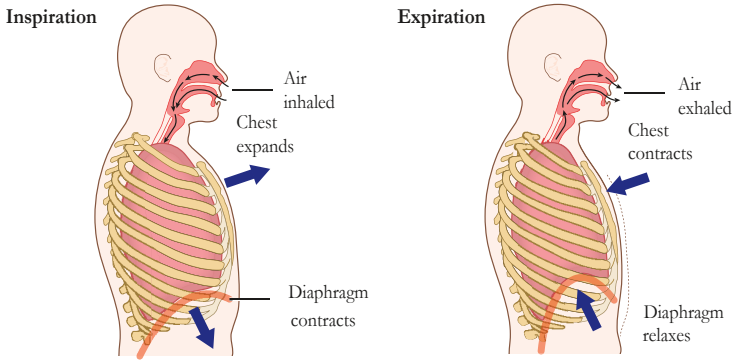


Figure 1.1: The two phases of breathing; adapted from [4].

dependent on age. In Table 1.1 examples of the normal RR for some age groups can be found, indicated in breaths per minute (BPM).

1.1.1 Respiratory patterns and apnea

Abnormal respiratory patterns, e. g. with largely varying RR or with cessations of breathing (COBs), may require clinical observation or intervention, or may indicate underlying conditions. When the RR is higher than the normal range, the term tachypnea is used, whereas lower RRs correspond to a bradypnea. Changes in the depth of breathing, i. e. deep and shallow, are referred to as hyperpnea and hypopnea. Apnea is used to identify the absence of breathing. Two main families of respiratory disorders are characterized by the occurrence of apneas, sleep-disordered breathing (SDB) and apnea of prematurity (AOP). These are considered as the two targeted applications for this work. Most of the contributions included in this thesis will focus on infants. However, the solutions proposed can be potentially generalized for both settings, taking into account the differences (e. g. environments, population, monitoring devices). Therefore, this chapter will introduce and discuss both SDB and AOP.

Table 1.1: Average RR at rest by age [5].

Age	**Average resting RR**
Birth to 6 weeks	30-60 BPM
6 months	25-40 BPM
3 years	20-30 BPM
6 years	18-25 BPM
10 years	15-20 BPM
Adults	12-20 BPM

SDB includes abnormal breathing patterns during sleep, such as apnea, hypopnea, and Cheyne–Stokes respiration. An apnea can be defined as a COB of at least 10 s [6]. Hypopneas are characterized by shallow breathing for more than 10 s with ventilation reduced by 30% to 90% and the presence of at least 4% reduction of oxygen saturation [7]. Cheyne-Stokes is a complex breathing pattern, consisting of periodic alternating phases of apnea and breathing. The latter is characterized by a crescendo-decrescendo pattern [5]. Apneas are divided into two main categories, one is called obstructive apnea (OA), which is caused by a collapse of the upper airway, the other is central apnea (CA), caused by the cessation of respiratory effort or drive [7]. The most common form for patients with SDB is the obstructive sleep apnea (OSA). Moderate to severe OSA was reported to occur more frequently (17%) in men in the age group of 50-70 years, and is less common in women in the same age group (9%) [6]. It can also be present in children and the elderly. Central sleep apnea (CSA) is less common and occurs in between 5% and 10% of patients with SDB, and is frequently a consequence of a vascular disease [8]. Apnea can also manifest in a mixed pattern, i. e. mixture of obstructive and central cases [9]. OSA has been associated with many comorbidities. If left undiagnosed and untreated it can lead to increased risks of hypertension, stroke, heart failure, diabetes, car accidents, and depression [10], and typical consequences of living with SDB are fatigue, tiredness, and lack of energy [11].

In premature infants, AOP is a typical respiratory disorder. In particular, AOP is present in nearly all infants born before 29 weeks gestation or with a birth weight (BW) below 1 kg, 54% of the infants with a gestational age (GA) between 30 and 31, and lower for higher GAs [12]. It is linked to the immaturity of the respiratory control system. Therefore, AOP commonly resolves with maturation, and it is usually classified as a developmental disorder [13]. Thus, there is an inverse correlation between the incidence of AOP and the GA and BW. The definition of apnea in the infant case is a COB that lasts at least 10 s if accompanied by bradycardia or desaturation, or a COB that lasts 20 s. Different apnea types can be present, central, obstructive, and mixed. Examples are visible in Figure 1.2. The most common apnea type in premature infants is mixed apnea (MA), approximately 50% of long apneic episodes (minimum duration 15 s) [15,16].

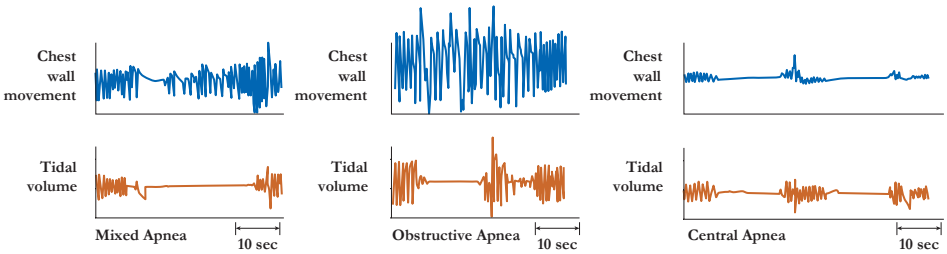


Figure 1.2: Apnea types: mixed, obstructive, and central; adapted from [14].

In CA cases respiratory flow stops due to the absence of breathing effort, whereas OA is caused by the collapse of the upper airway, particularly the pharynx [13,17]. Periodic breathing (PB) is a breathing pattern distinct from AOP, it can occur in term and preterm infants, usually for higher postmenstrual age (PMA). It is considered to be benign and is characterized by an alteration of COBs and breathing. The breathing pauses in PB can last up to 10 s [18]. In infants, long apneic events can lead to hypoxic-ischemic injuries of the brain. Moreover, studies are still establishing the presence of links between apneas and neurodevelopmental impairment, like cerebral palsy [12].

1.2 Respiration monitoring

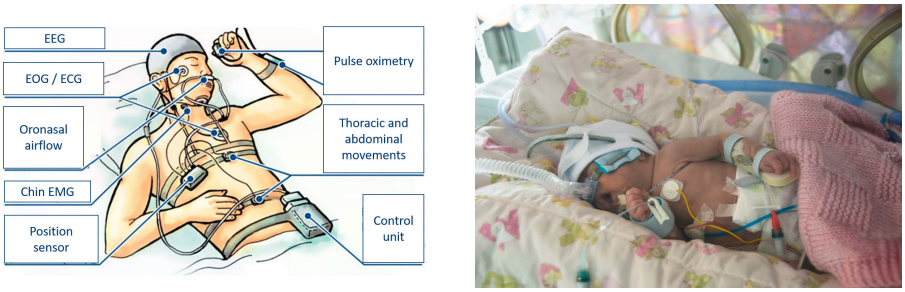
Monitoring and diagnosing breathing disorders is essential. One of the main vital signs that needs to be monitored continuously is the RR, next to heart rate (HR) and oxygen saturation. RR changes and respiratory irregularities, particularly apneas, need to be detected. Respiration can be monitored by monitoring the respiratory flow or the respiratory motion. Indirect methods to detect respiration have also been used. The commonly used methods in clinical practice require, usually, contact with the body or positioning sensors in the facial areas. Contactless solutions have also been developed to reduce patients' discomfort.

1.2.1 Traditional methods

SDB, and in particular apneas, can be diagnosed and monitored using polysomnography (PSG), where two respiratory motion bands and flow sensors are used to monitor both respiratory flow and motion [19]. This enables accurate apnea and hypopnea detection. Figure 1.3a shows the sensors and electrodes commonly used during PSG. In neonatal intensive care units (NICUs), chest impedance (CI) is usually used as respiration can be monitored using the electrodes already used to measure the electrocardiogram (ECG). A photo of an infant cared for in a NICU is visible in Figure 1.3b, the photo clearly shows the current invasiveness of monitoring. The methods available to monitor respiratory flow and motion are detailed in the following sections.

Respiratory flow

The gold standard method to monitor respiratory flow is the pneumotachograph which can be used to collect the temporal trend of inhaled and exhaled air and quantify the airflow. The air needs to flow through a resistive component which will cause pressure variations measured through a pressure transducer. The necessity to use a sealed face mask to ensure the lack of airflow leaks makes this method not well tolerated by patients, and particularly by infants [22, 23]. Temperature-



(a) **(b)**

Figure 1.3: Examples of current monitoring **(a)** graphical representation of a subject undergoing a PSG study [20]; **(b)** photo of an infant in a NICU [21].

based sensing can be also used to exploit the difference in temperature between the warm exhaled air and the inhaled one from the environment. Different sensors can be used, e. g. thermocouples and thermistors. These sensors have usually a small dimension and can be integrated into facial masks or positioned close to the nostrils and/or mouth to detect temperature variations, usually at the upper lip. The airflow measure obtained with these methods is not considered as accurate as the pneumotachograph, being a more qualitative estimation than a quantitative one [24]. Moreover, a clear limitation is the displacement of the sensors [25]. Intranasal pressure transducers have been commonly used in combination or as an alternative to temperature-based sensors due to their increased sensitivity [26]. They are less inconvenient than pneumotachograph but require the use of a nasal cannula and present limitations in the case of patients who are mostly mouth-breathers [24,27]. Hot wire anemometers are used similarly to pneumotachographs and exploit the cooling effect produced by the air flowing on a heated wire, the sensors can be embedded in a ventilator line. The main drawback lays in the sensor fragility [28].

Another difference between inhaled and exhaled air is the concentration of carbon dioxide. Capnography exploits this difference to monitor respiratory flow, also this measure is considered qualitative. Different sensors can be used to achieve this, the most common one is based on infrared spectroscopy, which exploits the absorbance peak of carbon dioxide at $4.26\,\mu m$ [29]. While this measure is available in most ventilators, it is mostly used in intubated patients in critical care and anesthesiology, for verifying the endotracheal tube placement, and the cardiopulmonary resuscitation effectiveness [30].

Respiratory motion

Breathing mechanics can be monitored using different sensors. Respiratory belts are very common in different environments. They consist of belts positioned

around the abdomen and/or the thorax of the subject measuring variations in the circumference. Different working principles are possible. Respiratory inductance plethysmography (RIP) uses two elastic bands, one for the thorax and one for the abdomen, the bands are equipped with coils. Changes in the inductance are measured reflecting the respiratory motion [24]. The sum of the abdominal and thoracic signals, after proper calibration, provides an estimation of the tidal volume [27]. Alternatively, the sensing can be performed using different technologies, e. g. strain gauges or fiber optics [22, 31]. Accelerometers, gyroscopes, or their combination can be also used to monitor chest movements [32].

The use of impedance pneumography or CI, is very common in clinical care, especially in NICUs [23]. The advantage of the method is the possibility of using the same electrodes used to measure the ECG, eliminating the need for additional sensors. The technology works by measuring changes in the impedance when injecting a low high-frequency current. The changes in air volume due to the respiratory activity are reflected in changes in impedance [32]. The tetrapolar solution is more accurate but would require the use of additional electrodes, therefore, usually, the bipolar solution is used [22]. Diaphragmatic electromyography (EMG) is also a possible solution to monitor the diaphragm's activity, the less invasive solution uses surface electrodes [33]. However, the method shares limitations with CI, as motion and cardiac artifacts [27], and obtaining a pure signal is not trivial due to cross-contamination [34].

Indirect

Respiration activity can be also monitored through cardiac activity, due to the effect the first has on the latter. Several modulations have been identified. This effect can be exploited to extract a respiration signal from the ECG and photoplethysmography (PPG) signals [31]. Algorithmic solutions are being used to fuse amplitude, baseline, and frequency modulations aiming at obtaining an accurate respiration signal [35]. The clear advantage of these methods would be obtaining two signals with the same sensing technology. However, questions about accuracy and differences between different patient populations are still present [36].

Apnea detection and alarms

The detection of apneas is linked to the accurate detection of a respiratory waveform. Generally, methods that monitor respiratory flow tend to be uncomfortable to wear but can detect any type of apnea. However, they are unable to distinguish between the different types. On the other hand, methods that monitor respiratory motion are generally able to detect only the occurrence of CAs, as the other two types can still present respiratory effort. In clinical environments, apnea detection is performed using specialized equipment like patient monitors or PSG.

Focusing on SDB, PSG is considered a gold standard diagnostic test [37] thanks to the many signals monitored. Sleep stages, hypopneas, and apneas are detected and scored by a polysomnographic technologist to determine the apnea-hypopnea index (AHI). The main problem, in this case, is associated with the discomfort of the patients undergoing PSG studies, and the fact that PSG recordings need to be analyzed and scored by trained polysomnographic technologists, making this solution a costly technique not available to all patients [37]. Home sleep apnea testing (HSAT) is used as an alternative for home monitoring of apnea occurrence in subjects who do not have other sleep disorders. Sleep staging is not possible due to the reduced range of signals monitored and limitations are present which could lead to false negative results or underestimation of the severity of OSA [19].

In the NICU environment it is clinically important to detect apneas. First of all, as long apnea events can have serious consequences, direct intervention is often needed to stop the event, e. g. through tactile stimulation [38]. Furthermore, monitoring the number of apneas, their type, and the severity of the events is relevant to decide on the maintenance or discontinuation of therapies, e. g. caffeine and/or respiratory support. Finally, the presence of apneas is relevant when deciding to discharge infants, as infants may only leave the hospital if they are considered apnea-free [39]. Therefore, alarms are generated if an apnea event lasting longer than a certain amount of time is detected by the patient monitor. This threshold is often 20 s [40].

CI is used in NICUs to obtain a respiration signal. Two typical problems in this environment are the generation of false alarms and missing apnea events [41, 42]. The first can be caused by poor electrode contact, motion artifacts, and shallow breathing [43]. Missing apnea events, on the other hand, can be caused by two main problems. CI is able to monitor respiratory effort, therefore the occurrence of OAs and MAs cannot be detected. The other problem that can cause missing an apnea event is the presence of a cardiac artifact [23]. Cardiac artifacts are a typical problem of CI, that in the absence of respiration detects the cardiac activity, during bradycardia the HR is in the range of normal breathing for infants [43, 44]. Therefore, in the clinic, nurses can experience a loss of confidence in the apnea alarms due to the false positives, apnea alarms were indeed reported to be one of the least reliable according to NICU nurses [45]. Considering the clustering of apnea, bradycardia, and desaturation (ABD event) [46], a combination of these alarms is often implicitly used by nurses to decide whether a clinically relevant apnea is occurring. However, it has been observed that in the case of very long apnea events (duration higher than 60 s) HR and oxygen saturation fall later and more slowly compared to the shorter ABD events [47]. Consequently, if the patient monitor fails in detecting the apnea event, the clinical intervention required can be significantly delayed. This can have serious consequences on infants and can result in prolonging apnea events when the event could have been interrupted if accurately detected and followed by a prompt intervention [47].

1.2.2 Unobtrusive alternatives

All methods presented until now require positioning electrodes or sensors on the face or body, the use of facial masks or nasal prongs, or wearing thoracic/abdominal belts. These solutions are perceived as extremely uncomfortable for patients, especially when combined with sensors to monitor other vitals [48]. Moreover, in NICUs the use of contact solutions can cause skin irritation and injury to premature infants [49, 50]. Furthermore, the high number of sensors and wires resulted in impeding parent-infant bonding [21]. Several contactless alternatives are being developed to reduce discomfort and to enable unobtrusive continuous respiration monitoring and apnea detection.

Most of the unobtrusive methods are able to monitor respiratory motion. Among these, radar has been widely researched due to the high distance of usage and ability to penetrate non-metal materials. Radars can monitor chest-wall movements, thanks to the Doppler effect, obtaining both respiratory and cardiac activities. Developments on this technology are focusing on both algorithms and hardware [51]. While several works analyze the performance on healthy subjects in lab settings [52–57], recent research has started moving to the clinical environment as well [58–62].

Pressure-sensitive films are also being integrated into bedding and mattresses. As radars, they can monitor both respiratory and cardiac activities by detecting mechanical vibrations. The sensors, usually piezoelectric or electromechanical, can be an array and output the distribution of pressure in the area, which requires strategies to combine or select array elements, or output a single signal, which can be heavily polluted by any motion artifacts. Several studies have been performed [63–65] some in the clinic [66, 67]. Capacitive electrodes can be integrated into the bedding as well, to monitor respiration activity similarly to CI [68]. Motion artifacts and baseline wandering are common problems [69].

Cameras

Cameras have also been used to monitor respiration. They have the main advantage of providing a video, i. e. images of the patient, which also delivers contextual information. In particular, red-green-blue (RGB) and near-infrared (NIR) cameras can monitor respiratory motion. RGB cameras or monochrome cameras with sensitivity in the visible range have been widely used for this purpose [70–79]. However, poor light conditions in most clinical environments combined with the necessity of continuous monitoring, i. e. overnight as well, limit the future use of this technology. Many solutions do not use make use of color information making the algorithms easily transferable to videos collected using NIR cameras, which represent a more viable alternative. When used in combination with NIR illumination, these cameras can monitor respiration also in complete darkness without disturbing the subject [80–84]. However, the required additional illumination complicates the measurement setup and increases the cost.

Recently, cameras with depth sensing are also acquiring more interest [85–89]. While the advantages of adding the depth information would fall on an additional detector for respiratory motion, it should be considered that it may share limitations with simple camera solutions, e. g. both would not visualize respiratory motion if a blanket is positioned to hide it. All the contactless methods presented until this point can monitor respiratory motion but not flow, therefore, even if their possible future use in clinical practice would reduce the obtrusiveness of the monitoring, it will anyway maintain limitations in the detection of apneas that have obstructive characteristics.

Thermal cameras detect infrared radiation naturally emitted from objects. Cameras sensitive in the range of mid-wave infrared (MWIR), 3 to 5 μm, and long-wave infrared (LWIR), 8 to 14 μm, provide adequate sensitivity for medical use [90]. They have been used in different health fields, but thanks to the independence of light conditions combined with the ability of monitoring both respiratory motion and flow, they gained interest in respiration monitoring. Studies with cameras sensitive in the MWIR were initially conducted [91–93]. Afterwards, LWIR became more popular, several studies were conducted in lab conditions [94–99], in outdoor environments [100], and in the clinic [101–103]. Works that combine thermal images with other camera modalities have also been proposed [104–106]. Thermal cameras are traditionally costly, however, low-cost solutions are recently being investigated for respiration monitoring [100, 105, 106]. Thermopile arrays are also low-cost thermal devices, they have been used for several applications, e. g. fever screening and fall detection [107, 108], but their use for vital signs monitoring has not been investigated yet. Overall, multi-thermal-camera setups would be required when aiming at continuous respiratory motion and flow monitoring. Therefore, the use of low-cost thermal solutions would allow the implementation of multi-camera setups without reaching prohibitive costs.

Background on video processing for respiration monitoring

Research has been focusing on different challenges related to camera-based respiration monitoring. Generally, a common sequence of processing steps can be identified in studies that aim at detecting respiration from videos. Firstly, the part of the images that contain the respiration information needs to be identified. This step can be referred to as region of interest (ROI) detection. Several strategies have been proposed to achieve this. Many works use manual selection of ROIs sometimes combined with tracking [70, 99–101, 109–113]. Other strategies involve the detection of specific landmarks, e. g. nose (for thermal imaging) or chest/abdomen detection [83, 91, 95, 97, 99, 104, 106, 114]. Skin detection has been also proposed [73, 79]. However, these solutions can result in being difficult to generalize for all sleeping positions and are further complicated for infants. Therefore, methods based on a specific feature or a combination of features were developed as well, e. g. pseudo-periodicity, [71, 75–77, 102].

The ROI detection step is commonly followed by strategies to extract the respiration signal. This is very frequently obtained from the average of the pixels' intensity or based on motion estimation, e. g. optical flow (OF). The use of geometrical variations of the ROI was also proposed [79]. Moreover, indirect methods based on remote PPG are also possible [113]. Steps that focus on artifacts reduction and detection can be considered, artifacts may be caused by motion or obstructions. For example, strategies aiming at motion and clinical intervention detection were studied [75, 77, 79]. Finally, the last step involves the analysis of the respiration signal to detect significant parameters or events. This analysis can be performed retrospectively, or by providing information with low latency. RR is commonly calculated, and apnea occurrences have been also detected [73, 99, 106].

1.3 Problem statement and objectives

In this section, we first describe the challenges that will be addressed, and later the objectives of this thesis.

1.3.1 Problem statement

Contact-based methods to monitor respiration require the use of belts, facial masks, or adhesives, causing discomfort or even skin irritations to patients. Camera-based respiration monitoring has been researched as an unobtrusive alternative to traditional methods, nevertheless, its application in hospital environments still requires further research and development. This thesis explores the possibility of using video respiration monitoring in a clinical setting and proposes technical and algorithmic solutions aiming at continuous respiration monitoring and apnea detection.

Particularly, we identified several problems that need further attention before such technology could be implemented in hospital environments. Firstly, the possibility of detecting apneas and COBs with low latency from the respiration waveform obtained from videos needs to be addressed. Secondly, algorithmic strategies to locate the pixels containing respiratory flow and respiratory motion independent of facial and body landmarks need further development. Moreover, with the purpose of detecting and identifying apneas, respiratory flow should be detected separately from respiratory motion in thermal recordings. Finally, a problem common to most technologies is the sensitivity to motion artifacts. Camera-based solutions are not immune to this problem. Strategies aiming at motion artifact rejection and robustness are needed to explore the possibility of reliably and continuously using this technology in real settings.

1.3.2 Objectives

We aim at developing a camera setup and algorithms to improve and extend the usage of cameras for respiration monitoring and apnea detection in the clinic. Based on the problems listed above, we define the objectives of this thesis.

Obj. 1: the development of a strategy to detect apneas from a respiration waveform. The method should decide on the presence of apneas with low latency opposed to methods that work in a retrospective manner. Furthermore, the possibility to detect short COBs should also be taken into account, as these may be relevant for apnea prediction purposes.

Obj. 2: the design of a setup using cameras low-cost enough to enable a multi-camera setup. Thermopile arrays' usability and their limitations for respiration monitoring need to be explored.

Obj. 3: the introduction of a multi-camera setup in the clinic for the collection of videos. The recordings will enable the development of algorithms and the analysis of working conditions in clinical settings.

Obj. 4: the development of a reliable method to automatically identify the pixels containing respiration information overcoming the limitations associated with facial/body landmarks detection.

Obj. 5: the analysis of the motion patterns and occurrence of movements during the recordings. This will allow estimating of the possible coverage of our system and developing strategies to improve the motion robustness and reject severe or problematic movements.

Obj. 6: the automatic separation of respiratory flow from respiratory motion in thermal videos where both signals are present. This has the potential to improve apneas detection and classification.

1.4 Thesis outline

This section illustrates the content of each chapter of this thesis. The chapters are briefly described and the associated publications are indicated.

Chapter 2 presents a method for the detection of COBs and apneas with low latency. The method has been developed and tested on respiration signals extracted from infant's videos collected with RGB and NIR cameras. The method is based on the comparison of a short-term and a long-term standard deviation, which allows monitoring amplitude changes in the respiration waveform. The recordings contained several short COBs which are used to demonstrate and estimate the performance of our solution. Finally, a method from literature,

that works retrospectively, is used as the benchmark. This contribution has been published in [C1].

Chapter 3 analyses the usability of a thermopile array for the detection of respiratory flow as an alternative to microbolometer-based devices. This is a feasibility study and videos were collected on healthy adult volunteers in lab settings. Several conditions are considered, variable distances, orientations, different respiratory patterns, and RRs. An initial strategy to automatically select the pixels containing respiration is also proposed. This contribution has been published in [J1].

Chapter 4 proposes a method to combine multiple camera views and an algorithm to automatically identify the pixels containing respiration. This algorithm is based on the combination of three features and it is independent of facial/body landmarks detection and visibility. Videos acquired in a neonatal ward using three low-resolution thermal cameras are used. Only the moments in which the infants are not moving are considered in this study. Furthermore, the method is compared to a similar method from literature. This contribution has been published in [J2].

Chapter 5 extends the contribution of Chapter 4 by combining the respiration monitoring algorithm with motion detection and classification. This allows increasing the usability of the solution moving towards continuous unobtrusive respiration monitoring. The videos acquired in a neonatal ward using three low-resolution thermal cameras are used. The motion in the videos is quantified and if considered severe by our classification an indication that the infant is moving is provided, otherwise, the respiration and RR are indicated. This contribution has been published in [J3].

Chapter 6 presents an algorithm to automatically distinguish the pixels containing respiratory flow from the ones containing respiratory motion in thermal videos. Aiming not only at respiration monitoring but also at improving apnea detection and classification, the differentiation of respiratory flow from motion is necessary. The algorithm is based on features that allow an automatic selection of the respiratory flow pixels independent of facial landmarks detection. Moreover, the positions of the respiratory flow in the thermal videos acquired in a neonatal ward are also analyzed. The contribution presented in Chapter 2 is used in this work as a proof of concept to demonstrate the advantages of monitoring respiratory flow and motion when OAs occur. This contribution has been published in [J4].

Chapter 7 analyses the long-term societal impact of this thesis under different points of view, i. e. the patients and their families, the clinicians, and the industry. This chapter contains also a description of the knowledge utilization referring to two patent applications filed during this work, [P1] and [P2].

Chapter 8 summarizes the results and identifies future research directions.

CHAPTER 2

On-line cessations of breathing detection

Abstract

Apnea detection is essential in neonatal settings as hypoxia can lead to permanent impairment. Short cessations of breathing are widespread in infants and their monitoring could be useful for the prediction of longer apneas. The aim of this chapter is to investigate the accuracy of our on-line cessation of breathing detector. Signals obtained through camera-based respiration monitoring were analyzed in five infants with 91 annotated cessations of breathing. The method proposed is based on the comparison of short-term and long-term standard deviations allowing the detection of sudden amplitude reduction in the signal with low latency. A new strategy able to detect short cessations of breathing on-line was successfully validated yielding an average accuracy of 93%.

This chapter is based on:
I. Lorato, S. Stuijk, M. Meftah, W. Verkruijsse, and G. de Haan, "Camera-based on-line short cessation of breathing detection," in 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp. 1656–1663, IEEE, 2019.

2.1 Introduction

Vital signs are of critical value to check the health of premature infants. Their monitoring is, therefore, standard practice in neonatal intensive care units (NICUs). Since apnea of prematurity (AOP) is common in this population, continuous respiration monitoring is crucial [23]. Apneas are prolonged pauses in the respiration and are common in infants with a gestational age (GA) below 34 weeks [115], the hypoxia typically associated with apnea could cause long-term or permanent impairment [116]. Respiration monitoring based on chest impedance (CI) currently used in NICUs presents limitations when detecting apneas, in particular, cardiac artifacts are a common cause of missed apnea detection [43]. Moreover, motion artifacts and other thoracic movements can also be misinterpreted as respiration [117]. Apneas are strictly defined as a cessation of breathing (COB) longer than 20 s or a COB of 10 s accompanied by bradycardia and/or desaturation [118]. However, discussions on the definition of clinically relevant apneas move the focus also on shorter COBs [119]. Short apneic episodes are common in infants and are defined as a respiratory pause of at least 3 s [120]. Moreover, this type of events can provide insights on the infant's respiratory system [121] and lead possibly to apnea prediction [122].

Since adhesive electrodes and sensors can cause stress or even skin damage to infants' sensitive skin [123], research in this field has been focusing on alternative non-contact respiration monitoring techniques. Among these, radars [124, 125], red-green-blue (RGB) or near-infrared (NIR) cameras [70, 72, 126], vision system based on depth sensing [127,128], thermal cameras [101,111], and pressure-sensitive films [66, 129] are the most researched for respiration monitoring in a NICU environment. Cameras represent one of the best solutions for NICUs applications. In the first place, because they are completely unobtrusive passive sensors and they allow to monitor multiple vital signs simultaneously. Moreover, cameras also provide contextual information that would be useful to nursing staff for infants' observation, and it would promote family-centred NICUs through live video feed to parents [130].

In this chapter, an approach for the on-line detection of short apneic events through camera-based respiration monitoring is proposed. Many methods for respiration detection algorithms using cameras have been developed targeting an infant population. Jorge et al. [131] proposed a camera-based approach for respiration monitoring based on a skin detection algorithm, which is not ideal for NICUs applications since infants' abdominal areas are commonly covered with blankets or snuggles. Though approaches based on remote photoplethysmography (PPG), e. g. [113], also rely on skin visibility, they can work on facial skin that is more likely uncovered. However, motion robustness necessitates multi-wavelength cameras. RGB cameras are a seemingly logical choice, but visible illumination may disturb sleep and is, therefore, not allowed. Proposals using wavelengths in the infrared range suffer from high cost of multi-wavelength cameras, or parallax

when using 3 cameras in parallel. Therefore, remote PPG-based solutions are not straightforward in such a complex environment. However, when monitoring respiration based on motion, skin visibility and color information are not indispensable for the signal detection. For example, Allinovi et al. [76] proposed a method based on maximum likelihood modeling and motion magnification able to automatically select the region of interest (ROI). The method proposed was tested on a limited dataset of adults and infants videos, with a window size of 20 s for estimation of the respiration signal and the respiration rate (RR). The latency (caused by the processing window) is particularly important when aiming at apnea detection and therefore, the method proposed by Janssen et al. [77] was preferred as a starting point for our work. The method, called video respiration monitoring (VRM), was extensively tested on adults videos, but limited experiments were performed on infants. Still, we consider this method very appealing for NICU-applications, particularly because of its attractive automatic ROI detection independent of skin visibility and the low latency of the method.

The output of the VRM algorithm is used as starting point for the detection of short apneic events in the respiration signal. Other works have been focusing on apnea detection strategies starting from video extracted respiratory signals, Jorge et al. [73] proposed an approach based on camera where COBs longer than 20 s were classified based on the RRs. If the RRs of the videos were lower than 20 breaths per minute (BPM) for a period longer than 20 s and no other motion was present in the video segments then it was classified as an apnea. However, aiming at the detection of short apneas time-domain approaches are preferable, being more sensitive to particularly short variations. Also Cattani et al. [71] tested the detection of apneas on camera-based respiration signals. The apnea detection strategy consisted in comparing the time-domain signal with an empirical constant threshold equal to 0.14. Constant thresholds have the drawback of not being able to adapt dynamically to changes in the signals, such as reductions in amplitude. Lee et al. [43], instead, proposed an approach based on modeling the distribution of normal breathing patterns and apnea ones, reaching an average detection performance over 90% by analyzing CI signals. The approach is, however, suitable only for retrospective analysis as specified by the authors, since the empirical parameters were optimized after filtering and baseline removal of the entire signal. This method has been widely used in apnea related publications, e. g. [42, 47, 132, 133], and has also been employed for the detection of short apneic events [122], therefore we decided to use it for comparison purposes.

The main contribution of this chapter is the development of an on-line short COB detection strategy based on the comparison of a short-term standard deviation with a long-term standard deviation. The RR is obtained as a byproduct of our processing. The rest of the chapter is organized as follows: Section 2.2 explains the method used and the dataset, Section 2.3 presents the results. Sections 2.4 and 2.5 contain respectively the discussion and the conclusions.

2.2 Materials and methods

This section details the materials and methods used for the development and testing of the COB-detector.

2.2.1 Materials

The videos used in this chapter were collected in the neonatal ward of the Máxima Medical Centre (MMC) in Veldhoven, The Netherlands. Two different setups were used for the data collection. Both studies received approval from the MMC and one study also received approval from the Internal Committee for Biomedical Ethics in Philips Research (ICBE2013-41-3797). Informed parental consent was obtained for all infants involved in the studies.

Dataset

The videos were annotated by a single person, the COBs were annotated only when clearly visible in the video. In total, 5 infants were included, Table 2.1 shows the postmenstrual age (PMA) expressed as the GA plus the postnatal age (PA), the total duration of the videos per infant, and the number of short apneic event annotated. The dataset includes both videos containing COBs and videos not containing any cessation events for control purposes. The videos have different duration going from 1 minute to 5 minutes reaching a total cumulative duration of 190.4 minutes. In total 91 short apneic events were annotated, the average duration and standard deviation of the COB population are 5.4 ± 1.9 s.

Table 2.1: Video details and parameters of the infants in the dataset.

ID	PMA (weeks)	Number of videos	Total Duration (min)	COB annotated
1	36+6.71	10	16.5	11
2	30+4.85	20	34.3	4
3	30+2.42	10	49.8	31
4	30+2.42	10	46.2	24
5	29+1.14	9	43.6	21
Overall		59	190.4	91

Experimental setup

The dataset includes videos collected with two different setups. In both cases, the CI from the patient monitor (Philips MX800) was also acquired for reference purposes. In the first study, a camera (UI-2220SE, IDS) was positioned on a tripod to have a view of the infants' chest/abdomen area, and some videos were collected from the top and others from the side. The videos were collected under visible light conditions with a frame rate of 20 frames per second and with a resolution of 576×768 pixels. Since color information is not relevant for respiratory motion detection, we used grayscale images. The videos were selected based on the quality of the reference signal and on the light conditions since the dataset also included measurements taken in particularly dark settings. Two infants (ID 1 and 2 in Table 2.1) were selected with a total video duration of 50.8 minutes.

The second setup included a monochrome visible light camera with the NIR filter removed (UI-22330SE, IDS) positioned on the incubator using suction cup mounting and visualizing an overview of the infant. NIR custom-made illumination was used since the normal workflow of the NICU was not disrupted and the incubator was covered as common practice, limiting the ambient light. The illumination unit comprised of LED arrays at three different wavelengths (660, 760, and 810 nm). The illumination level of all LEDs resulted in being around 0.2 mW/cm2 at the skin level of a patient, below the imposed limits (ANSI/AAMI/IEC 60601-2-21:2009). The videos were collected with a frame rate of 15 frames per second, a resolution of 608×864 pixels, and subjected to compression. In this case, the videos were selected only based on the quality of the reference signal. Using this setup, three infants are recorded as part of this dataset with a total video duration of 139.6 minutes.

2.2.2 Methods

Figure 2.1 summarizes the principal steps of the processing algorithm. The NICUs videos are input to our processing. The VRM algorithm of Janssen et al. (detailed in the next section) is used to extract the respiratory signal. On this respiratory signal, we run our COB-detector. Additionally, we compute and output the RR. In our benchmarking, we shall compare our COB-detector with the results from Lee et al. [43], and the RR with the CI-reference. The proposed algorithm was implemented in MATLAB (MATLAB 2018b, The MathWorks Inc., Natick, MA, USA).

Video respiration monitoring algorithm

The VRM algorithm proposed by Janssen et al. in [77] is a respiratory motion detection algorithm based on optical flow (OF). The algorithm automatically detects the ROI for respiration detection and returns the respiration signal. When non-respiratory motion is detected, the respiration waveform is put to zero and

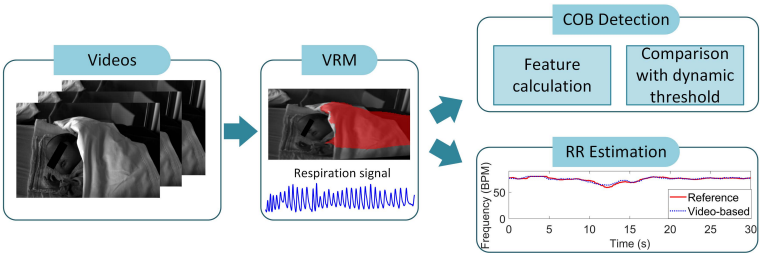


Figure 2.1: Main processing steps leading to the detection of COBs.

a template indicating that motion unrelated to respiration is present can be obtained. The same parameters introduced in the paper for the neonatal case [77] were used in this work.

The CI signal and the respiration signal obtained from the videos are both filtered using a bandpass Butterworth filter of the 4th order between 30 and 80 BPM since this is the normal range of RR in NICU infants including tachypnea cases [134]. The signals have different sampling frequencies corresponding to 15 frames per second or 20 frames per second depending on the acquisition and are processed with a sliding window approach with a window size of 3 s and a slide of 1 frame.

Cessation of breathing detection

In the case of a central apnea (CA) breathing cessation, a strong decrease in amplitude of the respiratory signal can be expected. Hence, our proposed COB-detector aims at signaling relative decreases in the standard deviation of the respiratory signal. Such a decrease can be recognized, by a short-term standard deviation σ_s becoming significantly smaller than a long-term standard deviation σ_l. Parameters in such an approach are the window-lengths for computation of the two standard deviations, and the threshold to define if a drop is "significant". Therefore, two window lengths are defined: a short window l_s in which a feature, corresponding to the short-term standard deviation, is estimated, and a long window l_l in which the long-term standard deviation is calculated. The calculation of σ_l is performed as the median of the previously evaluated σ_s. The median operation was preferred to the average for its robustness to outliers that can be present as sudden high signal amplitude due to undetected non-respiratory motion. The duration of l_s and l_l was chosen considering the length of the targeted COBs, which varies from 3 to 10 s. Moreover, the short window should contain at least a single period of respiration to be able to detect also the RR. Since the minimum RR expected is 30 BPM, l_s can be minimum 2 s. We arbitrarily decided to use l_s equal to 3 s. Furthermore, a too long l_l will cause the threshold to not adapt

dynamically to changes in the amplitude of the signal. On the other hand, a too short l_l will result in adapting also during apneic events. Therefore, as a compromise, l_l was chosen to be equal to 11 s.

More formally, let $resp(nT_s)$ be the time-domain signal after filtering obtained either from videos or from the reference, n depends on the current window and it is defined as $n = 0 + (j-1), 1 + (j-1), ..., (N-1) + (j-1)$. Where, j indicates the current window, the number of samples per window is $N = l_s/T_s$ and T_s is the sampling time. Then the short-term standard deviation is evaluated according to:

$$\sigma_s(j) = \sqrt{\frac{\sum_{n=0+(j-1)}^{(N-1)+(j-1)} (resp(nT_s) - \mu(j))^2}{N}}, \tag{2.1}$$

where, $\mu(j)$ is the average of $resp(nT_s)$. Thus, a value corresponding to the standard deviation of the time-domain signal will be obtained for each 3 s window. The long-term standard deviation will be evaluated on a window length l_l equal to 11 s, however to reduce the delay in the detection, σ_l is estimated with a fewer number of σ_s until $j > H$ with $H = l_l/T_s$:

$$\sigma_l(j) = \begin{cases} \underset{1 \leq k \leq j-1}{\text{median}} (\sigma_s(k)) & \text{if } j \leq H \\ \underset{j-H \leq k \leq j-1}{\text{median}} (\sigma_s(k)) & \text{otherwise.} \end{cases} \tag{2.2}$$

In each window $\sigma_s(j)$ and $\sigma_l(j)$ are compared. If the ratio between the two standard deviations results in being lower than 33%, the jth window is considered to contain a COB and a binary template, CD, is created as follows:

$$CD(j) = \begin{cases} 1 & \text{if } \sigma_s(j) \leq \sigma_l(j)/3 \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

$CD(j)$ indicates if the window j contains a COB. Figure 2.2 shows two examples, a signal containing COBs and one without.

Lee's method is applied on each video retrospectively on both CI and VRM signals. The method returns a probability of apnea, which is then converted to the weighted apnea duration as the area under the probability curve. The limit on the duration of the apneas detected, previously defined by Lee et al. as 5 s, is adjusted to this case making the smaller apnea detectable equal to 3 s, and obtaining, therefore, a second binary template for reference.

In the VRM respiration signal, since also motion information is available, cessations are not considered when motion unrelated to respiration is present and the standard deviation value $\sigma_s(j)$ for a jth window containing motion is not considered in the calculation of the $\sigma_l(j)$.

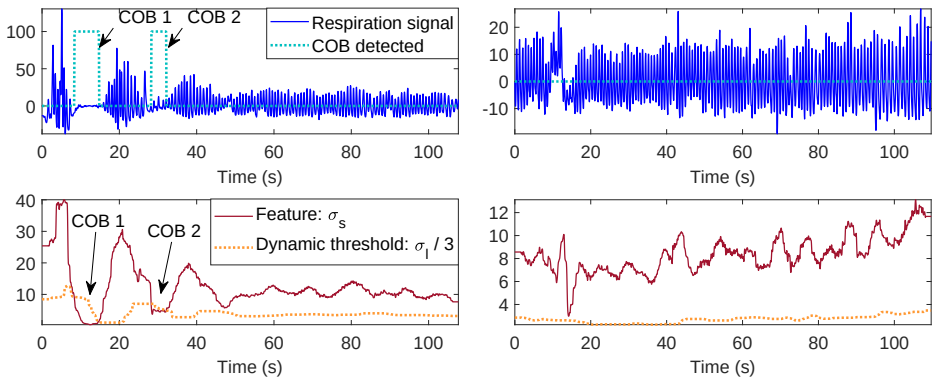


Figure 2.2: Examples of respiration signals, the two upper plots show a VRM respiration signal containing two COBs and a VRM respiration signal without cessations. The binary template, CD, labelled as "COB detected" has been multiplied for 100 for visualization purposes. The two plots in the bottom show the σ_s used as feature for COB detection and the threshold based on the σ_l.

Respiration rate estimation

In each 3 s window the RR is estimated as the frequency corresponding to the peak in the spectrum for both VRM signal and CI. The spectrum is evaluated using fast Fourier transform (FFT), zeropadding is performed reaching a number of samples equal to $120 \cdot N$. Moreover, to compensate for small variations, the RRs obtained are filtered using a moving mean filter followed by a moving median filter each with a window size of half a second.

Evaluation metrics

To compare RRs obtained with CI and VRM respiration signals the percentage of time in which the difference between the two is within ± 6 BPM is used as metric. This percentage has been evaluated in each video and then averaged. Moreover, for fair comparison, the RRs estimated in windows containing COBs according to the annotations and those estimated in windows where motion of the infant unrelated to the respiration was detected from the VRM algorithm have not been considered in the calculation of this metric.

To evaluate the difference between the COB detection algorithm proposed in this chapter and the one proposed by Lee et al. [43], sensitivity (SE) and specificity (SP) are calculated for each method using the manual annotation of the videos as reference. As defined in [71] SE will be:

$$SE = \frac{TTP}{TTP + TFN}, \tag{2.4}$$

with time true positive (TTP) and time false negative (TFN) being respectively the total duration of the time intervals with COBs detected correctly and with COBs incorrectly missed by the algorithm. And SP will be:

$$SP = \frac{TTN}{TTN + TFP}, \tag{2.5}$$

with time true negative (TTN) being the duration of the time intervals with no COBs in which there are no wrong detection, whereas time false positive (TFP) is the total length of the time segments with no COBs in which COBs are erroneously detected. Therefore, SE represents the ability of the algorithm to correctly detect COBs when present, whereas SP is the ability to correctly exclude the presence of COBs particularly important to avoid false alarms. The accuracy (ACC) can be, therefore, defined as:

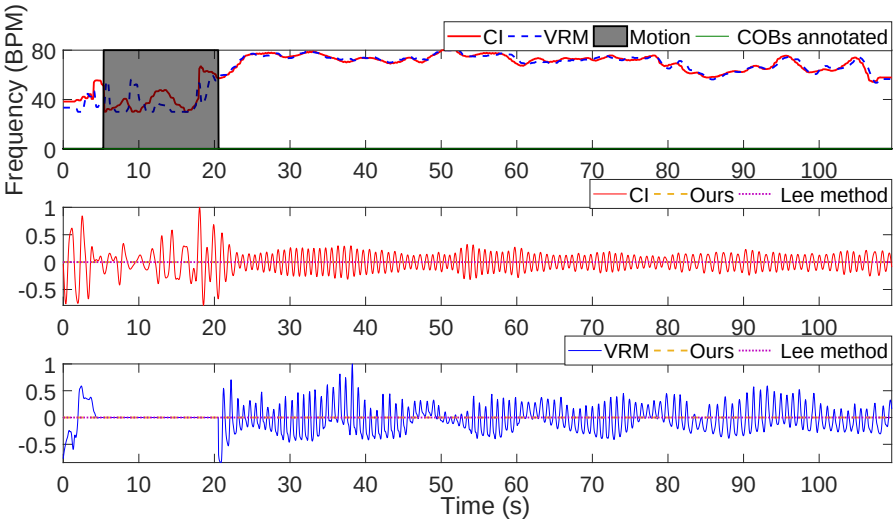
$$ACC = \frac{TTN + TTP}{TTN + TTP + TFP + TFN}. \tag{2.6}$$

2.3 Results

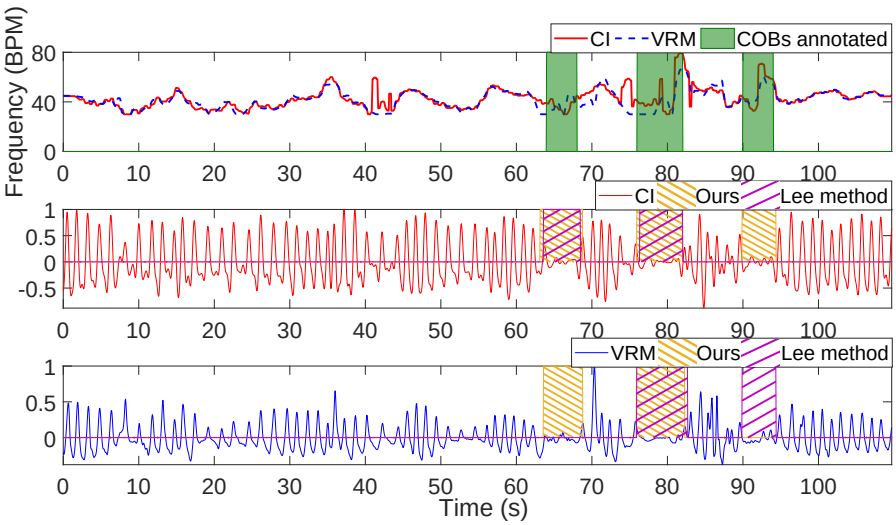
Table 2.2 contains the average results obtained for all the infants, sensitivity, specificity, and accuracy were estimated by comparing the detection of COBs performed with the method proposed in this chapter and Lee's method on both CI and VRM signals. In bold are indicated the best results comparing our method and Lee's. Figure 2.3 shows some examples of the obtained results. Figure 2.3a contains an example where motion but no COB was present, whereas Figure 2.3b shows the results obtained when three COBs were present. In the CI case, the COBs were all correctly detected by our method, whereas the detection is incomplete in the video signal case for both the proposed method and the benchmark one.

Table 2.2: Average SE, SP, and ACC results obtained with the method proposed in this chapter, indicated as ours, and Lee et al. method [43].

		SE (%)	**SP (%)**	**ACC (%)**
VRM	Ours	76.32	**94.39**	**93.16**
	Lee's method	**86.68**	91.50	90.64
CI	Ours	**83.15**	96.97	96.00
	Lee's method	77.02	**97.99**	**96.60**



(a)



(b)

Figure 2.3: Example of results obtained: (a) CI and VRM respiration signal and RR when motion unrelated to respiration is detected; (b) case with COB annotated. The amplitudes of the respiration time-domain signals were scaled to 1 for visualization.

2.4 Discussion

The method proposed in this chapter proved to be able to detect short apneic events. Preliminary results were obtained with videos of five infants containing

91 COBs in total. The comparison between our method and the one previously proposed by Lee et al. shows that our method resulted in higher specificity and accuracy for the respiration signals obtained from the videos, whereas the opposite happens in the CI case. Moreover, the results were always higher in the case of CI signal compared to the VRM signal using our COB-detector. This is most likely due to the noisiness of the signals, the VRM respiration signal relies on the correct detection of the ROI, that can be momentarily lost after strong movements causing low amplitude in the signal that can be misinterpreted as COBs. On average for both CI and VRM signals our method reached an accuracy of 94.6% against the 93.5% of Lee's method. Considering the high false alarm rate already present in the NICUs [45] it is of paramount importance to prefer specificity to sensitivity especially in the short apnea cases for which clinical relevance is still under discussion. Moreover, our method is able to work in an on-line fashion, whereas Lee et al. claim that their method requires several minutes of clean signal to work accurately and that an adaptation is needed for on-line detection [43], this is mostly due to the filtering and removal of baseline from the signal before the estimation of the moving standard deviation, on which the parameters of the Fermi function were optimized. Our method is also based on standard deviation but proved to work with a sliding window of 3 s, making COBs rapidly detectable by the system.

Moreover, the RRs obtained with the video signal and the CI as reference were compared. The RR extracted from the VRM signal is 75% of the time within 6 BPM from the CI one. Higher errors were obtained for infants with ID 3,4, and 5, this can be due to the compression of the videos and/or to a higher number of events with small motions unrelated to respiration. It should be considered that most of the studies using frequency-based RR detection used windows ranging from 8 to 20 s [73, 76, 113] in our case the 3 s window in which the FFT was performed leads to fast estimations but can also cause higher errors due to the poor frequency resolution.

This chapter introduces a new method for the detection of short apneas that can work on-line. The results are still considered preliminary, firstly the COBs were not annotated by an expert, however, the use of videos and not of CI for the annotations makes the result less subjective. Moreover, also longer apnea should be considered, there is no suggestion that such a method would not work for a different COB population, however, parameters such as l_l would need to be adjusted. The parameters used in this chapter were chosen arbitrarily or based on reasoning and the same parameters were chosen for CI and VRM signals, an optimization of these parameters could lead to improved results.

VRM delivered a respiration signal from videos with varying orientations and settings, e. g. containing motion or with infants covered. The algorithm is characterized by a set of empirically chosen parameters, adjusting the parameters could lead to improved respiration signals thanks to a more accurate detection of unwanted motion and a faster adaptation of the ROI following big movements.

2.5 Conclusions

The method described in this chapter proved to be able to detect short apneic events yielding an accuracy equal to 93.16% in the video signal case. The method can detect COBs with low latency and it is based on the comparison of short-term and long-term standard deviations. The detection of such short apneic events could lead to apnea prediction preventing hypoxic damages in infants. The method described in this chapter will be used in Chapter 6 to detect simulated obstructive apneas.

CHAPTER 3

Respiratory flow monitoring using a thermopile array

Abstract

Low-resolution thermal cameras have already been used in the detection of respiratory flow. However, microbolometer technology has a high production cost compared to thermopile arrays. In this chapter, the feasibility of using a thermopile array to detect respiratory flow is investigated in multiple settings. To prove the concept, we tested the detector on six healthy subjects. Our method automatically selects the region of interest by discriminating between sensor elements that output noise and flow-induced signals. The thermopile array yielded an average root mean square error (RMSE) of 1.59 breaths per minute. Parameters such as distance, respiration rate, orientation, and oral or nasal breathing resulted in being fundamental in the detection of respiratory flow. This chapter provides the proof of concept that low-cost thermopile arrays can be used to monitor respiratory flow in a lab setting and without the need for facial landmark detection. Further development could provide a more attractive alternative for the earlier bolometer-based proposals.

This chapter is based on:
I. Lorato, T. Bakkes, S. Stuijk, M. Meftah, and G. de Haan, "Unobtrusive respiratory flow monitoring using a thermopile array: a feasibility study," Applied Sciences, vol. 9, no. 12, pp. 1–15:2449, 2019.

3.1 Introduction

Continuous respiratory monitoring can be used to predict potentially serious clinical events like cardiopulmonary arrest [25, 135], and it is particularly relevant for the detection of specific clinical conditions such as apnea of prematurity (AOP), sleep-disordered breathing (SDB), or sudden infant death syndrome (SIDS) [97]. Respiratory activity can be monitored with a variety of techniques including flow sensors, chest wall motion sensors, and indirect extraction of the respiration signal [25]. However, in some applications, flow sensors are necessary. In particular, obstructive apnea (OA) is a type of apnea characterized by the presence of chest wall motion, but absence of flow [23]. Various sensors can be used to detect respiration flow such as pneumotachographs, oro-nasal thermistors, or nasal pressure transducers. These, however, need to be positioned close to the nostrils, fixed on the upper lip, or integrated in a mask and result in being particularly uncomfortable for patients [23].

Many remote sensing technologies have been used to monitor respiration, e. g. red-green-blue (RGB) and near-infrared (NIR) cameras or radars. Their detection is, however, based on respiration motion, not flow. An unobtrusive way to monitor respiration flow is based on thermal imaging, also called infrared thermography (IRT) [93, 136]. Most of the work on IRT uses the nostrils and/or the mouth area as the region of interest (ROI) to evaluate the thermal-induced respiratory variations. The ROI can be determined manually or automatically by using anatomical characteristics, e. g. medial canthus region, combined with tracking algorithms [91, 92, 101].

In applications such as sleep monitoring, multiple cameras would be necessary to monitor the patient continuously and cover all possible orientations. This would undoubtedly raise the cost of the solution, making it even more important to find an inexpensive sensor solution to detect respiration (and the absence of it). Cho et al. [100], for example, used a FLIR ONE to detect respiration flow in outdoor and indoor settings. The work focused on the robustness to environmental thermal dynamics. Scebba et al. [137] proposed a sensor fusion of FLIR Lepton and RGB cameras. Both works used ROIs focused on the nostrils area and relatively small microbolometer-based thermal cameras. Microbolometers are the most common sensors for uncooled infrared imaging: the infrared radiation absorbed changes the resistance of the material [138].

Thermopile arrays, instead, consist of multiple thermopile sensors. These sensors are constituted by thermocouples in series, which are based on the Seebeck effect, i. e. they generate a voltage proportional to the infrared radiation absorbed [139]. They have been widely used for non-contact thermometers [140], fall detection [107], human detection [141–143], and fever screening [108]. Thermal cameras based on microbolometers started having smaller dimensions and cost, as FLIR ONE or FLIR Lepton. However, the main difference between thermopiles and microbolometers lies in the production cost, microbolometer-based infrared

imagers require several costly processes that significantly raise the price [144]. It would, therefore, be of interest to understand if a thermopile array could be used as a respiratory flow detector, yielding a further reduction in cost. Moreover, the detection of facial landmarks is not straightforward in more complex environments [102], especially when the resolution would be further reduced. A recent work from Pereira et al. [102] proved the usability of a thermal camera to detect the respiration signal without the use of anatomical features, both on healthy adults and infants in the neonatal intensive care unit (NICU). The method uses a black-box approach, wherein the resolution is reduced by averaging multiple pixels. Moreover, an empirical signal quality index (SQI) is used to analyze the spectrum of each cell, and the ROI containing respiratory information is chosen based on the SQI value. In the data used, both face and torso are included in the field of view, as specified in a previous work where the same dataset was used [95]. The respiration signal detected is a combination of motion-related breathing (due to shoulder motion) and thermal changes due to respiration flow. The proposed approach eliminates the need of ROI selection based on facial features. However, an unconstrained combination of motion and flow-related information is not desirable and should not be trusted in environments where apneas and, in particular, OAs are common.

In this chapter, we present an unobtrusive respiration monitoring technique that uses a thermopile array and a ROI detection method that does not require facial landmarks' visibility. The accuracy of the breathing signal detection was assessed in optimal stationary conditions on healthy adults in a laboratory environment as a proof of concept. To prove the feasibility of detecting respiration with a thermopile array for typical respiration conditions, oral/nasal breathing and different respiration rates (RRs) have been analyzed. Furthermore, several other factors like distance from the sensor and orientation have been considered to mimic real-world applications.

The rest of the chapter is organized as follows, Section 3.2 introduces the background and the main challenges, and contains the processing steps and the experimental setup. The results and discussion will be presented, respectively, in Section 3.3 and Section 3.4. Finally, Section 3.5 provides the conclusions.

3.2 Materials and methods

Modern thermal cameras have a sensitivity in the order of 10^{-2} K [138]. However, a typical thermopile array can only provide a sensitivity in the order of 10^{-1} K. Considering that the thermal variations in both adults and neonates have been reported to be higher than tenths of a Kelvin [145], such a sensitivity should be theoretically sufficient for the detection of flow. The detection of respiration flow through thermal sensing is not only dependent on the amplitude of the thermal variation. Sensor characteristics also contribute in determining the ability to

detect such changes. The partial area effect [146] is a known dependency that describes the relationship between the real amplitude (temperature in this case), field of view, number of pixels, distance, size of the target, and the amplitude (temperature) detected by a sensor. It follows that the target size, i. e. the surface involved in thermal heat exchange, is an important parameter in the detection of flow. For example, differences are expected if the measurements are taken in frontal view, i. e. subject facing the device, or side view, with the head on a pillow. In the second case, the pillow material will also take part in the heat exchange process, increasing the area involved. This phenomenon was already observed by Abbas et al. [101]. The partial area effect is particularly important in sensors with a broad field of view and a low number of pixels and limits the distance of usage.

Another challenge can be found in the RR, higher RRs will result in lower temperature variations due to the heat capacity. Moreover, oral and nasal respiration show differences in temperature and humidity content depending on the environment. In particular, oral exhaled air results in having a higher temperature compared to the nasal one with an ambient temperature of 20°C [147].

A single thermal sensing element equipped with proper optics would be enough for detecting the respiration flow. However, aiming such a sensor in the correct direction would result in being particularly complicated. Taking into account all the above considerations, a Panasonic Grid-EYE was selected to prove the feasibility of detecting respiration flow. The sensor has a sensitivity of 0.16 K and a resolution of 8×8 pixels with an angle of view of 60°, and it is particularly low-cost compared to other sensors on the market.

3.2.1 Materials

The thermal videos used in this chapter were collected on healthy volunteers using the thermopile array Panasonic Grid-EYE. The study was approved by the Internal Committee Biomedical Experiments of Philips Research with ID: ICBE-2-26849, and informed consent was obtained from each subject prior to the study.

Datasets

The videos were collected in different orientations, at different distances, and with different breathing patterns. The subjects were asked to lie down as still as possible on a bed or sit on a chair and to breathe spontaneously or according to an audio stimulus. Two different datasets were collected: the first one to prove the detection of respiration with such a sensor in different settings, and the second one to prove the feasibility in the detection of respiration flow and to assess the performance of the proposed algorithm. Each video had a duration of 2 min.

Dataset A: constant guided breathing in different settings
The videos were collected on a single subject since the purpose of the analysis performed was not to highlight inter-subject variability. For each condition, 3

videos were collected. Unless specified otherwise, the videos were collected during nasal respiration and with the subject lying down. Overall, Dataset A consists of 39 videos.

- Distance: videos were collected in side view with different distances, i. e. from 10 to 50 cm, and with a constant RR of 20 breaths per minute (BPM).
- Respiration rate: different constant RRs were used as the breathing pattern, i. e. from 10 to 30 BPM. Videos were collected in side view with a constant distance of 20 cm.
- Orientation and oral/nasal respiration: videos with a constant RR of 20 BPM and distance of 20 cm were collected in side and frontal orientations during both nasal and oral breathing.
- Shoulder motion: the videos were collected in a seated position, with a constant RR of 20 BPM and a variable distance, i. e. 50 and 100 cm.

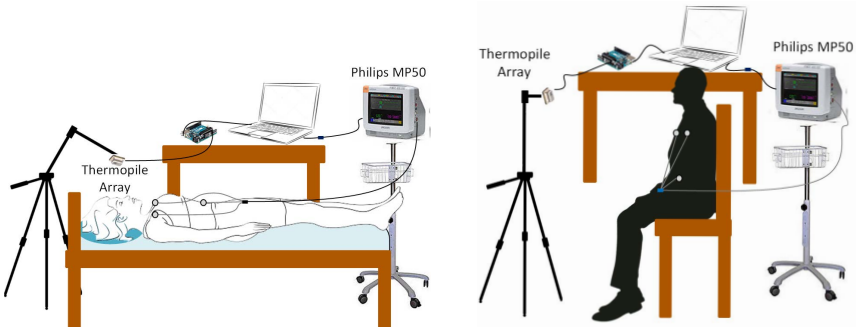
Dataset B: guided breathing with different patterns
Since the aim of this chapter is to provide a proof of concept, we chose to use 5 subjects similar to the number of subjects used in [100] for their proof of concept study. Videos were collected with guided and spontaneous breathing. For each subject, two videos at constant RR of 10 and 30 BPM were collected, a video with a linearly-variable RR, and a video with an RR with two sudden change from 10 to 30 BPM. Moreover, videos during spontaneous respiration were collected. All the videos were collected in side view during nasal respiration at a distance of 20 cm. Overall Dataset B consists of 25 videos.

Experimental setup

The thermal images were collected using a Grid–EYE thermopile array AM8854 from Panasonic. The sensor has 64 thermopile elements arranged in an 8×8 grid, and it is sensitive in the long-wave infrared (LWIR) range. The field of view covered by the Grid–EYE is 60°, and it has a frame rate of 10 Hz. The sensor was connected to an Arduino UNO through the I^2C bus, and the acquisition was performed through MATLAB.

The data collection was performed in a room with an ambient temperature of around 20°C. It should be specified that respiratory activity causes also motion of the head and shoulders. This kind of motion is visible also in thermal images. Therefore, measurements to prove feasibility in flow detection have been taken with a subject lying down and shoulders out of the field of view. Moreover, measurements in seated positions have also been collected to prove the detection of respiration based on motion. Figure 3.1 shows the experimental setup.

Chest impedance (CI) was used as ground truth for respiration monitoring. Since the measurements were acquired in a stationary condition, CI can be considered reliable. The signal has been acquired using three electrocardiogram (ECG)



(a) **(b)**

Figure 3.1: Experimental setup: (**a**) subject lying down on a bed; (**b**) subject in a seated position.

electrodes, as is common in clinical practice, and a patient monitor (Philips MP50) with a sampling frequency of 62.5 Hz.

3.2.2 Methods

The main steps of the processing chain are summarized in Figure 3.2. The videos are first preprocessed. Afterwards, a pixel containing the respiration needs to be selected, and the RR is extracted in the frequency-domain. The processing chain was implemented in MATLAB (MATLAB 2018b, The MathWorks Inc., Natick, MA, USA), and the data processing was performed offline.

Preprocessing

The collected data suffered from irregular sampling. Hence, in the first processing step, the data were interpolated to produce a dataset that was uniformly sampled in the temporal dimension with a sampling frequency of 10 Hz. Afterwards, the spatial mean was removed to suppress the thermal variations of the environment.

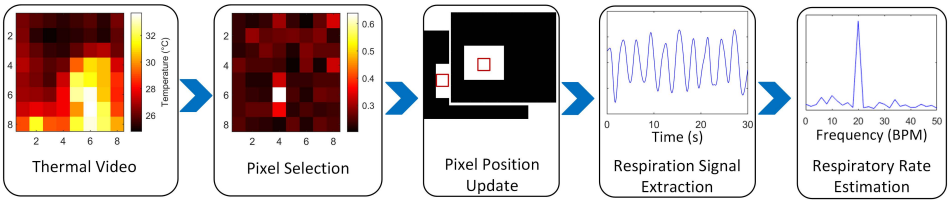


Figure 3.2: Block diagram summarizing the processing chain.

The videos were then processed with a sliding window approach with a window size of 30 s. The first RR estimate would, therefore, be available after 30 s and then updated. This time lag was considered acceptable since in some clinical environments manual estimation of RR is typically performed by nurses by counting chest movements over a period of time that goes from 15 to 60 s [148]. Moreover, such a window was also chosen taking into consideration the trade-off between frequency resolution and temporal resolution, typical of spectrogram methods. Using a time window of 30 s allowed obtaining a relatively good frequency resolution equal to 2 BPM. For each window, the pixels' signals were filtered using a Butterworth bandpass filter going from 10 to 40 BPM, since the normal RR in healthy adults lies in this range [113].

Pixel selection and pixel update

The pixel containing the respiratory information should be identified. An approach was proposed in [102] where an empirical metric, SQI, was evaluated based on four features of the spectrum. The SQI is dependent, however, on empirically chosen thresholds. The method to identify the pixel with the strongest respiratory signal proposed in this chapter also operates in the Fourier domain. For each pixel, the time-domain samples were Hanning-windowed, then transformed through a 1D fast Fourier transform (FFT), and the value of the highest peak in the normalized magnitude spectrum was used as a strength indicator. Consequently, an 8×8 matrix of strength indicators was obtained, and the maximum element in this matrix was used to identify the pixel containing the respiratory signal.

Formally, let $\hat{x}_{m,l}(nT_s)$ be the filtered, Hanning-windowed time-domain signal for the jth window where the spatial mean has been removed. m and l are used to indicate a pixel in the 8×8 images and $n = 0 + (j-1)$, $1 + (j-1)$, ..., $(N-1) + (j-1)$ depends on the current window j. The number of samples per window was $N = 300$, and the sampling time was $T_s = 0.1$ s. $y_{m,l}(f_k)$ is the spectrum of the preprocessed time-domain signal $\hat{x}_{m,l}(nT_s)$ and $\mid y_{m,l}(f_k) \mid$ its magnitude, with $f_k = k/(NT_s)$ Hz and $k = 0, 1, ..., N/2 - 1$. We define the height of the normalized spectrum's peak as:

$$q_{m,l} = \frac{\max\limits_{lim_1 \leq f_k \leq lim_2} (\mid y_{m,l}(f_k) \mid)}{\sqrt{\sum\limits_{f_k = lim_1}^{lim_2} \mid y_{m,l}(f_k) \mid^2}}, \tag{3.1}$$

where $lim_1 = 0.167$ Hz and $lim_2 = 0.667$ Hz, i. e. boundaries of the respiration band.

We assumed the respiration signal to be pseudo-periodic, thus, the spectrum would contain a peak. Noise, on the other hand, would show a relatively flat spectrum, as illustrated in Figure 3.3. In our algorithm, we interpreted a higher peak in the normalized spectrum as an increased likelihood that the pixel would

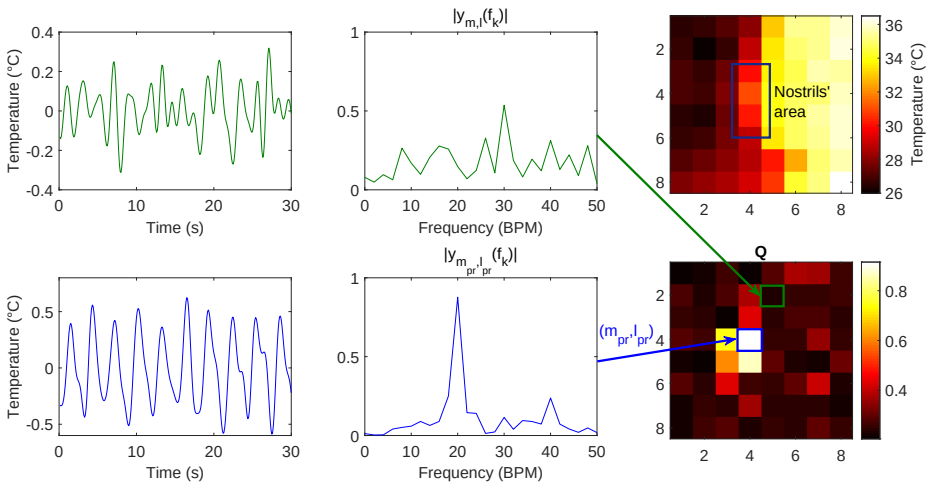


Figure 3.3: The upper plots show the time-domain signal and spectrum of a pixel where no respiratory information is present. The plots on the second row show the time-domain and spectrum of another pixel that contains a respiration signal. The upper figure shows a thermal image where a subject is in a profile view at 20 cm from the sensor, the lower one shows the distribution of the normalized spectrum's peak calculated per each pixel, i. e. **Q**. The pixels containing the signals plotted are indicated.

contain the respiratory signal, and we only used a single pixel in the array in the position (m_{p_r}, l_{p_r}) defined as:

$$(m_{p_r}, l_{p_r}) = \arg\max_{(m,l)}(q_{m,l}). \tag{3.2}$$

The pixel needs to be selected for each window j. For the first z windows, the position was chosen by selecting the pixel with the highest peak in the normalized spectrum. The successive positions were, instead, dependent on the preceding z positions, limiting the search space of the following pixels. This procedure was intended to prevent rapid pixel-switching, which could deteriorate the computed spectra, whereas the scenario (substantially static subject) did not require a rapid adaptation. A mask was created based on the z preceding positions that allowed the following ones to move only to an adjacent position. The value z was fixed empirically to 10 windows.

Respiration signal extraction and respiration rate estimation

Once the pixel is selected, the respiration signal is obtained for each window as:

$$resp(nT_s) = \hat{x}_{(m_{p_r}, l_{p_r})_j}(nT_s), \tag{3.3}$$

where $(m_{p_r}, l_{p_r})_j$ indicates the pixel selected with the proposed method in the jth window.

To estimate the RR, the spectral density (SD) can be calculated. The SD can be estimated as the squared magnitude of the Fourier spectrum, known as the periodogram method:

$$SD_{PER}(f_k) = y_{m_{p_r},l_{p_r}}(f_k) \cdot y_{m_{p_r},l_{p_r}}(f_k)^* = \mid y_{m_{p_r},l_{p_r}}(f_k) \mid^2 . \tag{3.4}$$

Several other methods have been developed to estimate a more precise SD, autoregressive (AR) models have already been used to extract respiration from photoplethysmography (PPG) and have shown good potential [149]. AR models are parametric methods for signal modeling and SD estimation. The model assumes that a signal can be modeled based on the weighted sum of p precedent values. In the spectral domain, in particular, the SD can be estimated as:

$$SD_{AR}(f_q) = \frac{\sigma^2 T_s}{\left| 1 + \sum_{b=1}^{p} a_b exp(-j2\pi f_q b T_s) \right|^2}, \tag{3.5}$$

where σ^2 is the variance of the noise, the a_b are the weights of the AR model, and p is the model order [150]. The weights need to be estimated and the order indicated. Moreover, it should be noted that this parametric model allows the estimation of SD_{AR} in an arbitrary number of points between 0 and $1/(2T_s)$. Therefore, in general, f_q can be different from the previously-defined f_k. AR models are not directly affected by the discrete Fourier transform (DFT) limitations because they estimate the frequency response directly, without implicit windowing and spectral resolution restrictions. Therefore, AR models have been used especially when spectra need to be estimated on short windows. This parametric method is suited for narrowband signals with medium or high signal to noise ratios (SNRs) [151].

Several methods are available for AR modeling, detailed information can be found in [152]. Burg's method was chosen for its stability and accuracy on short data recordings. The model order p is not known a priori, and since there is no optimal method yet, many methods have been proposed in the years to correctly estimate it [153]. Chaichulee et al. [154] recently proposed to determine the order by choosing the p that minimizes the error obtained by comparing the peak of AR's SD and the Fourier spectrum. In this chapter, Chaichulee's method has been used, the optimal model order was selected for each window, and f_q was chosen to be equal to f_k. In Dataset B, the RRs have been estimated with both the AR model and periodogram, to compare the performances. The range of order was empirically chosen and went from 15 to 30 with a step of 1 whereas in Dataset A the periodogram method was used.

The RR was estimated as the frequency corresponding to the highest peak. Once the RRs were estimated, a moving median filter was applied averaging 10 estimations together. Moreover, the time-domain signal was reconstructed by gluing together the segments, as explained in [155].

Evaluation metrics

SNR, RMSE, and relative error (RE) have been used to evaluate the performance of the thermopile in detecting the respiration signal. The SNR was evaluated according to:

$$SNR = 10\,log_{10}\left(\frac{\sum_{f_k=0.12}^{1.67}(U(f_k)\,SD(f_k))}{\sum_{f_k=0.12}^{1.67}((1-U(f_k))\;SD(f_k))}\right), \tag{3.6}$$

where $SD(f_k)$ is the signal's spectral density estimated using the periodogram or AR method and $U(f_k)$ is a binary template window centered on the spectrum's peak and first harmonic with a predefined width equal to 2 frequency bins. The first harmonic was included as the respiration signal is not a perfect sinusoid. The SNR was evaluated on each window, and the average of all the windows was used. The RMSE has been evaluated using CI as the reference:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N_w}\left(\widetilde{rf_j} - rf_j\right)^2}{N_w}}, \tag{3.7}$$

where $\widetilde{rf_j}$ is the RR estimated, rf_j is the one obtained using the CI, and N_w is the number of windows evaluated. RE has also been estimated to allow a better comparison.

Moreover, the videos collected in Dataset A have also been compared by assessing the number of pixels involved in the respiration signal. This was performed by evaluating the absolute value of the correlation coefficient between the detected respiration signal and the signals contained in other pixels. If the correlation was higher than 0.7, then the pixel was considered to contain a respiration signal as well and was added to the pixel count.

The temperature variation detected by the thermopile due to respiration was also considered. This was obtained by using the range of the signal, i. e. the maximum amplitude minus the minimum amplitude. Both the amplitude and the number of pixels have been estimated on each window and then averaged for the entire measurement.

3.3 Results

The results obtained using the thermopile array are presented for the two different datasets.

Dataset A

An overview of the analysis performed on Dataset A is shown in Figure 3.4. Multiple distances, RRs, and frontal or side nasal/oral breathing have been compared. The dots and squares represent the average of the three signals obtained for each setting, whereas the error bars show the standard deviations. Signals at distances of 20 and 30 cm, and 40 and 50 cm were grouped together, as a variation of 10 cm was not considered enough to show the effect of increasing distance. Therefore, in the distance case, the average and standard deviation were evaluated on six signals per setting. Moreover, Figure 3.5 shows the pixels containing the respiration signal in the case of a measurement taken at 50 cm in side view with the subject lying down and 50 cm in the frontal seated position, as well as the results obtained when the respiration was detected through shoulder motion.

Dataset B

Table 3.1 contains the average of the results obtained for each subject for the four different respiration patterns used, with both the periodogram and AR model. The last row shows the overall average results, in bold we indicate the best results for RMSE, RE, and SNR. Figures 3.6 and 3.7, instead, show the time-domain signals, the spectra evaluated through FFT on each window, and the RRs estimated in comparison with the reference for the five cases. It should be specified that in the case of guided breathing with a sudden change, an additional filter was applied. As can be expected from the results obtained in Dataset A, breathing at 30 BPM will have a lower amplitude compared to the 10 BPM one. This effect caused a delay in the detection of the frequency change. Therefore, a ramp filter has been applied to each window of the sudden change signals to amplify high frequencies in the breathing range. Correlation plots and Bland–Altman analysis for all the subjects and breathing scenarios are displayed in Figure 3.8. The mean bias was 0.23 BPM with the limits of agreement being -5.05 and 5.51 BPM. Moreover, the results also showed a strong correlation between the CI estimated RR and the one estimated with the thermopile array, yielding a $\rho = 0.95$.

3.4 Discussion

The results obtained for the two datasets and the different challenges are discussed separately.

Dataset A

- Distance: the comparison of the performance between different distances proved that the thermopile would lose accuracy at larger distances. In particular, amplitude, SNR, and the number of pixels would decrease with

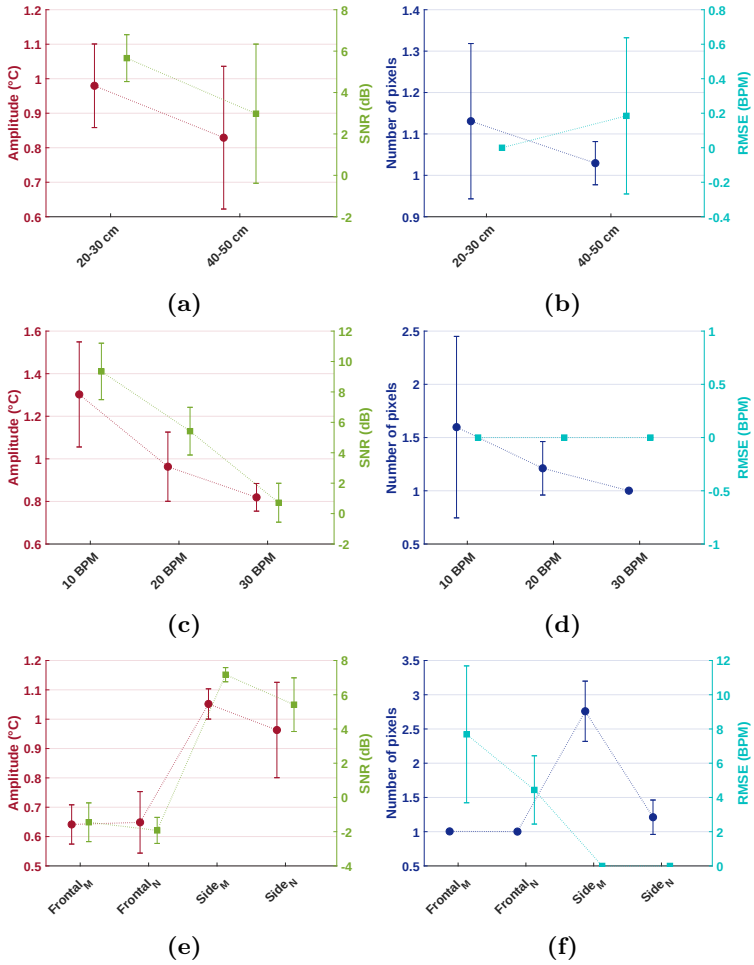


Figure 3.4: Overview of the differences in amplitude, SNR, number of pixels, and RMSE in Dataset A: (**a**,**b**) results for signals acquired at different distances; (**c**,**d**) comparison of different RRs; (**e**,**f**) different orientations and oral and nasal respiration, M and N indicate respectively mouth and nose breathing.

larger distances, and the RMSE would increase. Amplitude decreased from an average of 0.98°C to 0.83°C. The SNR reduced from 5.66 dB to 2.97 dB, whereas RMSE went from 0 BPM to 0.18 BPM with a standard deviation of 0.45 BPM. The number of pixels did not result in a big change on average, but it can be noticed that the standard deviation for the 20–30 cm case was much higher compared to the 40–50 cm case. This reduced performance with larger distances is mainly due to the partial area effect. As already expected,

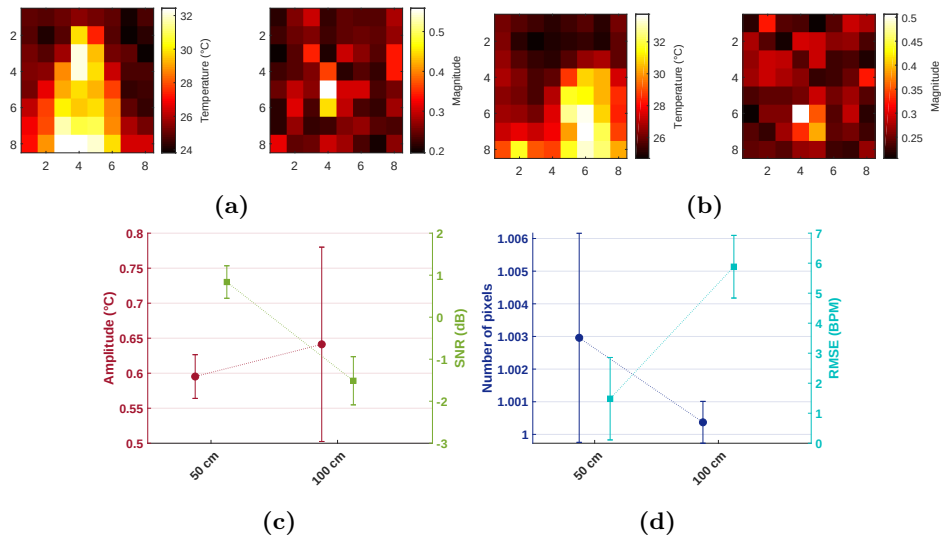


Figure 3.5: Respiration detection through shoulder motion: (**a**,**b**) the first figure shows a thermal image, and the second one shows the spectrum peak magnitude distribution of a subject seated and lying respectively at 50 cm of distance; (**c**,**d**) overview of the differences in amplitude, SNR, number of pixels, and RMSE when data were collected from a seated subject at 50 and 100 cm.

the low-resolution combined with the broad field of view limits the distance at which the sensor can be used.

- Increasing RR: breathing with higher RR resulted also in a reduction of the amplitude from 1.30°C to 0.82°C, and the SNR drastically reduced from 9.34 to 0.71 dB, whereas the number of pixels from 1.60 to 1.00. The RMSE

Table 3.1: Average of five subjects of the results obtained on Dataset B.

Breathing Pattern	Periodogram			AR		
	RMSE (BPM)	RE (%)	SNR (dB)	RMSE (BPM)	RE (%)	SNR (dB)
10 BPM	0.00	0.00	8.71	0.00	0.00	13.59
30 BPM	0.00	0.00	5.15	0.00	0.00	8.15
Chirp	1.41	1.19	4.09	1.41	1.21	5.87
Sudden Change	5.57	7.79	2.19	6.57	9.61	7.90
Spontaneous	0.98	1.59	4.27	0.99	1.72	5.81
Avg.	**1.59**	**2.11**	4.88	1.79	2.51	**8.26**

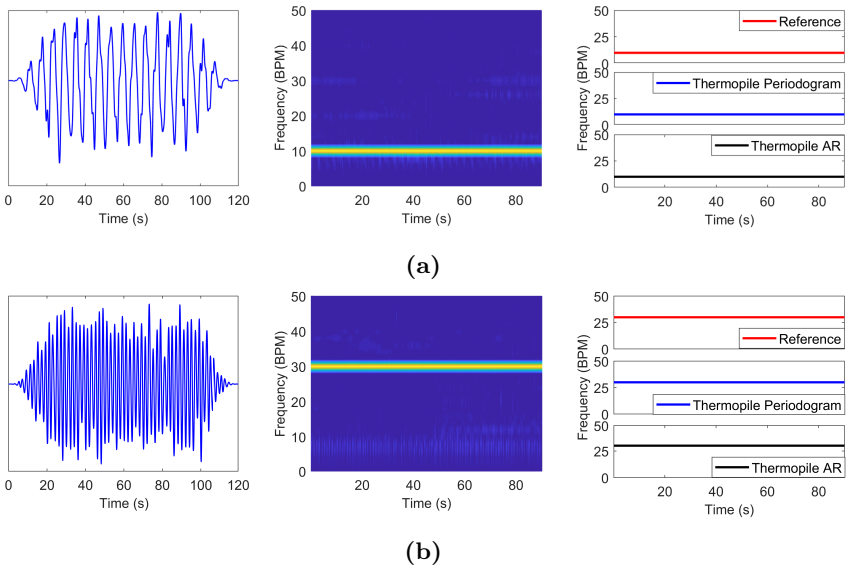


Figure 3.6: Results obtained for Dataset B with constant breathing pattern at 10 and 30 BPM, respectively; on the left the time-domain signal, in the middle the Fourier spectrum per each window, and on the right the RR extracted with both the periodogram and AR and the CI one.

remained constant. During higher RR the flow applied to the surroundings would be imposed for less time, reducing the heat exchanged and, therefore, the quality of the thermal signal.

- Orientation and oral/nasal respiration: in this dataset, both orientation and nasal/oral respiration were compared. Frontal measurements were expected to produce a reduced variation in temperature due to the small surface for heat exchange available. Indeed, both frontal nasal and frontal oral respiration resulted in performing particularly worse compared to the other settings. Frontal oral respiration performed worse compared to the nasal case resulting in a higher RMSE. Side oral respiration resulted in causing a bigger temperature variation and higher SNRs. This behavior was expected since in the side case, the pillow's textile material would be involved in heat exchanges as well, increasing the surface. This setting involved many more pixels, i. e. on average 2.76 compared to side nasal breathing, which resulted in 1.21 pixels on average. The variation in amplitude and SNR was not as big. The differences can be explained by considering the differences in expired air temperature as already indicated in Section 3.2 and differences in the flow direction.

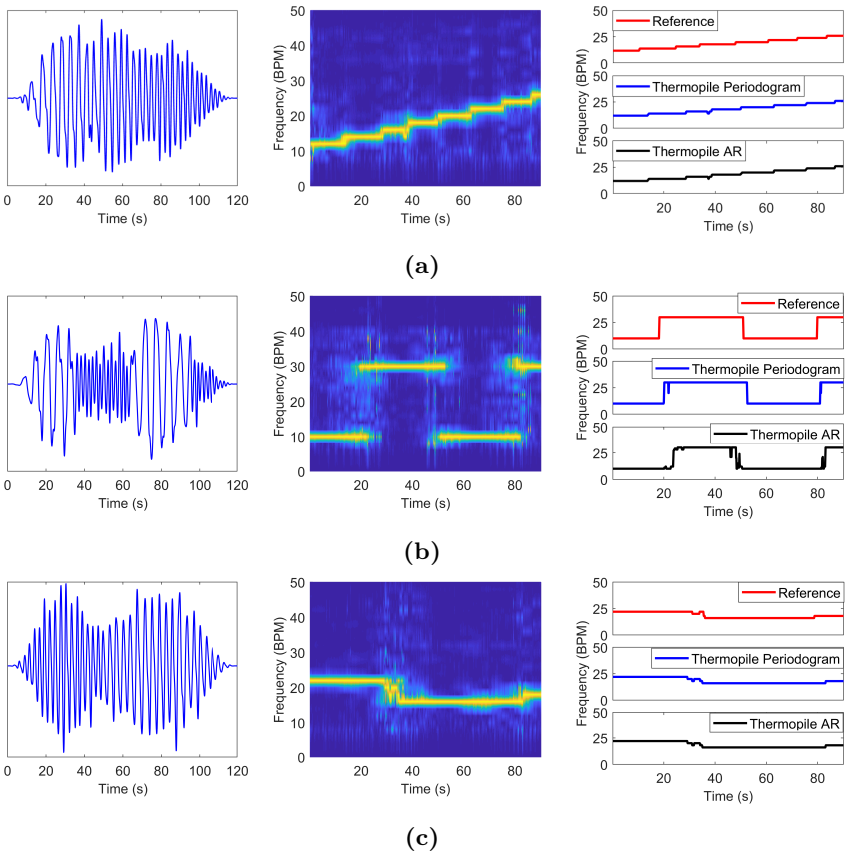


Figure 3.7: Results obtained for Dataset B with different patterns; on the left the time-domain signal, in the middle the Fourier spectrum per each window, and on the right the RR extracted with both the periodogram and AR and the CI one. (**a**) chirp pattern; (**b**) pattern with two sudden changes; (**c**) spontaneous breathing.

- Shoulder motion: measurements were collected in a seated position at 50 and 100 cm when the shoulders were included in the field of view. Figure 3.5 shows the two different situations: in the first case, the signal resulting in the maximum peak magnitude was coming from the shoulder area, whereas in the second case (side view with the subject lying down), the signal was coming from the nose area. This is caused by the higher amplitude of the temperature variation due to respiratory shoulder motion compared to the temperature variation produced by respiratory flow in the frontal view at the same distance. Shoulder and head motion due to respiration can be

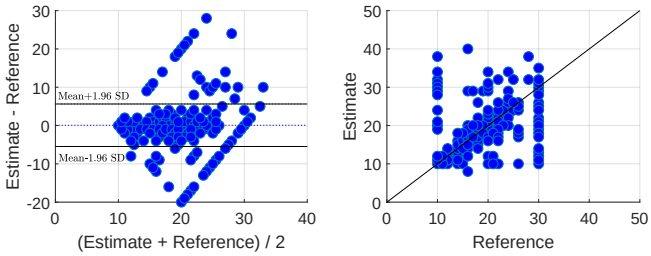


Figure 3.8: Bland–Altman plot and correlation plot for all the subjects and settings in Dataset B. Reference indicates the RR estimated with the CI, whereas estimate indicates the one estimated using the thermopile array.

suppressed by performing the measurement with the subject lying down, at least for adults. Moreover, due to the partial area effect, the error in the detection of respiration through motion increased drastically already at 100 cm.

Dataset B

Constant RR of 10 and 30 BPM both performed well, resulting in an RMSE of 0 BPM for both the periodogram and AR model. A difference in the SNR can be noted between the two RRs since the respiration at 30 BPM resulted in a lower SNR compared to the other one. This was already expected considering the results obtained in Dataset A. Chirp had a median error of 1.41 BPM for both the periodogram and AR model, whereas the pattern with sudden change resulted in the highest RMSE, 5.57 BPM and 6.57 BPM, respectively. This was expected since in the sudden change case when indecision between the two peaks was present (similar height of the peaks in the spectrum), an error of 20 BPM could be obtained. Moreover, the sudden change was the only case in which the AR performed particularly worse compared to the periodogram. This can be explained considering that the SNR in this case was significantly lower compared to the other four cases, and AR is known to perform well in cases of high or medium SNRs. In the spontaneous breathing case, the RMSE resulted in being around 0.98 for both SD estimators. Moreover, both correlation and Bland–Altman plots in Figure 3.8 show a bias in the estimated RR at 10 and 30 BPM. This bias was due to a delay in the detection of the transition between 10 and 30 BPM in the signals acquired with the sudden change pattern.

The periodogram yielded an average RMSE of 1.56 BPM compared to 1.79 BPM of the AR model, the RE was 2.11% and 2.51%, respectively, and the SNR resulted in being on average 4.88 dB and 8.26 dB, respectively. The difference in error was on average only 0.23 BPM, which we considered insignificant. On the other hand, the SNR was always higher in the AR model case. The similarity in the

results is justifiable considering that Chaichulee's method binds AR and Fourier spectrum evaluated through FFT to have the same RR estimation, therefore limiting AR models to the same limitations of DFT. AR modeling can improve the results obtained in short windows, but a new method for order selection should be investigated.

In this chapter, respiration detection using a thermopile array was proven to be feasible in controlled settings. It should be noted that this is a proof of concept, future research should focus on proving the usability of such a sensor with a higher number of subjects and in clinical practice. Some limitations can be identified in the current approach, such as a lack of motion robustness and the pseudo-periodicity assumption for the respiration signal. Strategies to compensate motion are required. Considering that RGB cameras can currently monitor multiple vital signs and would be probably included in clinical practice in the future, a combination of RGB videos and thermopile array images could be used to track the identified pixel. A similar approach has been proposed in [137] however, their solution relied on facial landmark detection. Moreover, the pseudo-periodicity assumption would not hold in cases of apneas or other respiratory patterns. Strategies to overcome this challenge are required before moving to clinical practice.

3.5 Conclusions

A thermopile array has been tested to prove the feasibility in detecting respiratory flow. The results showed that respiration motion dominates the measurement for a seated subject. This indicates that a lying adult subject is crucial to measure respiratory flow successfully, rather than head and shoulder motion due to respiration. Moreover, the results prove the sensor can detect respiratory flow in different ideal settings, and with variable RRs. The pixel selection method successfully detected the pixel containing the respiration flow thermal signature. The method for pixels selection was used as starting point for the development of the method presented in Chapter 4. The overall conclusion is that a thermopile array can be used for unobtrusive respiration monitoring. However, the usability has only been demonstrated in ideal conditions for healthy adults and extending the usage to a more challenging population, e. g. infants, and environment may not be straightforward. Further development may provide interesting low-cost alternatives to the earlier microbolometer-based proposals.

CHAPTER 4

Multi-camera infrared thermography and automatic respiratory pixels identification

Abstract

Respiration is monitored in neonatal wards using chest impedance (CI), which is obtrusive and can cause skin damage to the infants. Therefore, unobtrusive solutions based on infrared thermography are being investigated. This chapter proposes an algorithm to merge multiple thermal camera views and automatically detect the pixels containing respiration motion or flow using three features. The method was tested on 152 minutes of recordings acquired on seven infants. We performed a comparison with the CI's respiration rate yielding a mean absolute error equal to 2.07 breaths/min. Merging the three features resulted in reducing the dependency on the window size typical of spectrum-based features.

This chapter is based on:
I. Lorato, S. Stuijk, M. Meftah, D. Kommers, P. Andriessen, C. van Pul, and G. de Haan, "Multi-camera infrared thermography for infant respiration monitoring," Biomedical Optics Express, vol. 11, no. 9, pp. 4848–4861, 2020.

4.1 Introduction

Premature infants are cared for in neonatal intensive care units (NICUs) where their vital signs need to be continuously monitored to detect critical events. To measure these biosignals, like electrocardiogram (ECG), respiration, and oxygen saturation, many electrodes and sensors are applied on the infants' sensitive skin causing discomfort and, in some cases, also skin damage [156]. Respiratory frequency is usually monitored since respiratory instability and apneas, cessations of breathing (COBs), can be common in preterm infants and in term newborns with respiratory diseases [157], and may require immediate action of the caregiver. Chest impedance (CI) is used to monitor respiratory frequency in NICUs, which can be measured using the electrodes already applied for measuring the ECG. However, this method still suffers from motion artifacts and it is not reliable for neonatal apnea detection [43].

Unobtrusive or non-contact solutions to monitor respiration are being investigated for clinical environments. The two main observable effects of respiration are respiratory motion and respiratory flow. Respiratory motion can be monitored unobtrusively using different solutions as red-green-blue (RGB) or near-infrared (NIR) cameras [77, 79], radar-based solutions [57, 124], pressure-sensitive films [66, 158], or thermal cameras [95, 109]. Of these, thermal cameras are able to also deliver flow information unobtrusively [101, 159].

Respiratory monitoring is often used to detect apneas, which can be classified in three main categories: central, obstructive, and mixed. In central apneas (CAs) no stimulus for breathing is given by the brain and therefore, both respiratory motion and respiratory flow are absent, whereas in the other two categories some sort of respiratory motion, i. e. respiratory effort, is present [115]. Therefore, when aiming at apnea detection and classification, technologies that can detect both flow and motion are more appealing, like infrared thermography (IRT). Previous studies based on thermal cameras in a NICU environment used facial landmarks detection to locate the nasal area by exploiting for example the medial canthus region, one of the warmest areas of the face [91,97]. This is however quite complex in infants' thermal recordings and requires high resolution thermal images, or proximity of the camera to the nasal area [100], or a combination of thermal and RGB cameras [137]. Therefore, approaches aiming at automatic region of interest (ROI) detection in thermal imaging are being developed [102, 160].

Pereira et al. [102] designed an algorithm to automatically select the ROI containing respiration signals in infants. This algorithm relies on a large resolution reduction which allows reducing the noise in the thermal recordings. Afterwards, a signal quality index (SQI) is calculated for each ROI based on the spectrum and an empirical threshold is used to decide which ROI can potentially contain a respiration signal. Such an approach is very promising as it allows to obtain a respiration signal even when the nose is not clearly visible in infant thermal recordings. Our previous work [160], described in Chapter 3, proposed an alter-

native approach tested on a thermopile array, which is a very low-cost device that delivers low-resolution thermal images. The algorithm proposed for the automatic detection of the ROI was based on the height of the normalized spectrum peak and required no empirical thresholds to be defined. While the first work was developed for a high-resolution thermal camera and was tested also in infants, the second one was developed for an extremely low-resolution thermopile array but was tested only for adults in ideal conditions. However, both approaches propose the use of features for the localization of the pixels containing respiration based solely on the spectrum shape, this may be complex to generalize as different spectrum characteristics can affect the result. For example, the frequency resolution of the spectrum affects how the energy is distributed and if short windows are used, it may result in noise pixels' spectra being difficult to distinguish from respiration ones due to the spectral leakage, especially when the signal to noise ratio (SNR) is low.

Moreover, considering complex environments as NICUs, the use of multiple camera views is important to obtain proper coverage in all the infants' possible positions for the detection of both respiratory flow and motion signals, called in this chapter respiration thermal signature. Furthermore, since neonates in NICUs are inside incubators and the incubators' Plexiglas walls are not transparent for thermal cameras, these should be positioned inside the incubator which is not feasible for large high-resolution cameras. Even in the case of babies being in an open bed, bulky cameras can obstruct caregivers' and parents' interaction with the infant, therefore, the use of smaller thermal solutions should be investigated.

Building on our previous contribution [160], detailed in Chapter 3, this chapter proposes a new algorithm based on an automatic ROI detection for respiration monitoring in multi-camera low-resolution thermal videos. We propose a data fusion on a pixel level and define three features that can be merged together to obtain a more accurate localization of the respiration pixels without relying on spectrum shape only, spatial averaging, or facial landmarks detection. This algorithm was tested on thermal videos recorded on seven infants in a real neonatal ward, reaching a total amount of 152 minutes using CI as reference.

The rest of this chapter is organized as follows. Section 4.2 explains the method developed, the setup used, and the dataset. Section 4.3 presents the results. These are then discussed in Section 4.4 together with limitations and possible future research, and Section 4.5 contains the conclusion.

4.2 Materials and methods

The data collection performed, the videos collected, and the algorithm for the automatic respiratory pixels identification are described in this section.

Table 4.1: Infants' data, the gestational age (GA) is indicated in weeks (w) and days (d), postnatal age (PA) is in days

Infant	Gender	GA (w+d)	PA (days)	Weight (g)	Sleeping Position	Duration (minutes)
1	F	26w 4d	59	2445	Supine	34.52
2	M	38w 5d	3	4140	Supine	9.24
3	M	34w 1d	16	2995	Supine	31.17
4	M	26w 3d	59	2750	Prone	11.10
5	M	39w	2	3670	Lateral	10.76
6	F	40w 1d	6	3420	Supine	44.70
7	M	40w 2d	1	3775	Lateral	10.17

4.2.1 Materials

The study was conducted in the neonatal ward of the Máxima Medical Centre (MMC) in Veldhoven, The Netherlands and the study received a waiver from the Ethical Committee of the MMC. A written informed parental consent was obtained for all the infants included. The thermal videos were collected using three FLIR Lepton cameras.

Dataset

The inclusion criteria of the study were: (i) babies monitored using ECG electrodes for clinical purposes, (ii) clinically stable, (iii) in open bed. Both preterm and term babies were included in the study. In total fifteen babies can be included in this study. This chapter uses the recordings collected on the first nine babies, two babies have been excluded from further processing due to the blanket position that resulted in completely hiding the respiratory motion while the respiratory flow was also barely visible. Table 4.1 presents the information regarding the remaining seven infants. For each baby, around three hours of videos were acquired, except for Infant 7 where only one hour was acquired. Since we expected problems from interfering motion (examples baby moving, parents' or nurses' handling, or soother presence) we selected only moments without patient handling and patient motion. The videos were, therefore, manually annotated, the useful moments can vary from around nine minutes to forty-five depending on the infant. One baby with respiratory support was included for feasibility purposes and corresponds to Infant 1 in Table 4.1.

Experimental setup

The recordings were performed using three FLIR Lepton thermal cameras. These were Lepton 2.5, with a resolution of 60×80 pixels, the cameras are sensitive in

the long-wave infrared (LWIR) range, specifically from 8 to 14 μm. The thermal sensitivity is around 50 mK and the average frame rate is 8.7 Hz. The sensors were connected to smart I/O modules (Pure Thermal 2 or Pure Thermal Mini) and these were connected to a laptop through USB. The acquisition was performed using MATLAB, the temperature readings were obtained by using the cameras with the factory default calibration.

The cameras were positioned around the babies' open beds to cover all possible infant positions inside the bed, mounting arms as shown in Figure 4.1 were used. We chose to use multiple cameras to increase the camera coverage and have a good visualization of both respiration motion and flow considering different infants' positions. The two cameras pointing at the infant face from the two sides are called camera 1 and 2, whereas camera 3 is the camera on the foot side of the bed registering the entire baby area. We specify that camera 3 is actually a Lepton 3.5 which differs from the others only for the resolution, equal to 120×160. We noticed the 60×80 resolution was sufficient in detecting the respiration rate (RR) and decided to use all the cameras at the same resolution. The images of the Lepton 3.5 were, therefore, down-scaled to 60×80 simulating the use of a third Lepton 2.5. All three cameras have a shutter that has been deactivated during the acquisition to avoid delays caused by the closing and opening of the shutter. The reference signal, CI, was obtained from the patient monitor (Philips MX800) sampled at 62.5 Hz. An artifact was used to synchronize the videos with the patient monitor.

Figure 4.1: The three FLIR Lepton cameras positioned around an infant's open bed, two with white housing and one with blue housing.

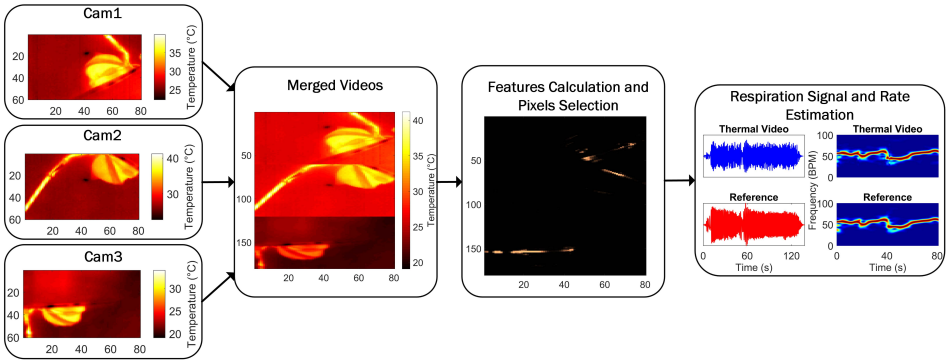


Figure 4.2: Major processing steps summarized.

4.2.2 Methods

The videos from the cameras are merged obtaining a single image plane. Based on three features a core-pixel, presumed to contain a strong respiratory signature, is selected. This core-pixel is then combined with temporally highly correlating pixels to form the respiratory output signal. Consequently, this output uses both respiratory flow and motion present in all camera views. The RR can then be estimated as the frequency corresponding to the peak of the spectrum. These steps are summarized in Figure 4.2 and explained in more detail in the next sections. Our algorithm will be benchmarked against Pereira's method [102], adapted to make it work with our hardware setup. The algorithm was developed and executed offline in MATLAB (MATLAB 2018a, The MathWorks Inc., Natick, MA, USA).

Preprocessing

The videos were recorded using MATLAB. However, due to the acquisition strategy, the sampling rate of the videos was not uniform. Therefore, to obtain uniformly sampled videos, a 1D linear interpolation, i. e. using the standard MATLAB function *interp1*, was performed for each pixel's time-domain signal resulting in three videos sampled at 9 Hz (the choice of the sampling frequency is not particularly critical and 9 Hz was chosen such that it was close to the average frame rate). At this point we considered two possible steps, one is treating the videos as three separate streams and then combine the RRs, the other one is combining the videos on a pixel level and process them as a single video. The first approach has some advantages based on the independence of each view, for example, if a movement is visible only in one of the cameras the other two can potentially still deliver a good RR estimation. The second one, instead, rejects weaker respiration signatures, that can be obtained from unfavorable camera views, allowing to use only the overall best pixels. Most importantly, however, the first

method does not perform well when only one of the cameras detects a correct RR, and therefore, the second method has been implemented. The result, after merging the videos, is a single video with spatial dimensions equal to $M \times L$ with $M = 180$ and $L = 80$ pixels as shown in Figure 4.2. It should be specified that we did not correct the single views' temperature values as visible in Figure 4.2 because the absolute temperature values do not affect the processing since each pixel's time-domain signal is used independently (the feature called *Gradient* is the only one that could be affected by differences in the temperature values as this could lead to a false edge-detection, however, the combination with the other features compensates for this effect).

The merged video was processed with a sliding window approach with a window size equal to 15 s and a slide of 1 s, the window size was chosen as a trade-off between accuracy and latency. A 15 s window results in a resolution of 4 breaths per minute (BPM), and the slide size results in an updated RR every second.

Automatic respiratory pixels detection

We based our respiration pixels selection strategy on different features that are used to find a core-pixel in the entire image, this pixel will be then used to find the other pixels that contain respiration. As the core-pixel is crucial to the result, the features are built to ensure a very strict selection. A simplified block diagram of the algorithm is shown in Figure 4.3. Each pixel is processed separately, let $x_{m,l}(nT_s)$ be each pixel's time-domain signal in a 15 s window and at the position (m, l). If we define the current 15 s time window as the jth window then n is defined as $n = 0 + (j-1)/T_s$, $1 + (j-1)/T_s$, ..., $(N-1) + (j-1)/T_s$ with N being equal to 135 samples, i. e. the samples in 15 s with a sampling time $T_s = 0.111$ s. Three features were used to select the core-pixel in each window:

- *Pseudo-periodicity*: the first feature is similar to the one presented in Chapter 3 and our previous work [160], i. e. the height of the normalized spectrum

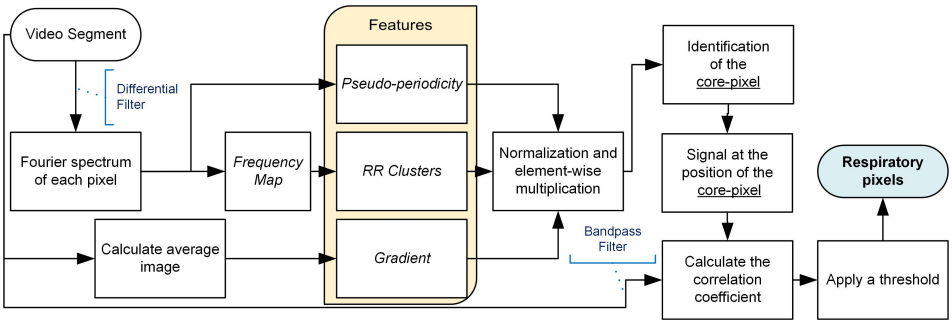


Figure 4.3: Simplified block diagram of the respiratory pixels selection algorithm.

peak that assumes the respiration signal to be pseudo-periodic and thus identifiable compared to noise. Each $x_{m,l}(nT_s)$ is filtered with a differential filter, the filtered version is named $x'_{m,l}(nT_s)$. This is transformed, after the application of Hanning window and zeropadded till $N_z = 120 \cdot N$, using a 1D discrete Fourier transform (DFT) obtaining $y'_{m,l}(f_k)$ with $k = 0, 1, ..., N_z/2 - 1$ and $f_k = k/(N_z T_s)$ Hz. and it is defined as

$$q_{m,l} = \frac{\max\limits_{0 \leq f_k \leq \frac{(N_z/2-1)}{N_z T_s}} (\mid y'_{m,l}(f_k) \mid)}{\sqrt{\sum\limits_{f_k=0}^{\frac{(N_z/2-1)}{N_z T_s}} \mid y'_{m,l}(f_k) \mid^2}}. \tag{4.1}$$

The $q_{m,l}$ calculated for each pixel are elements of the $M \times L$ matrix called $\mathbf{Q}$, see example in Figure 4.4. The feature defined in Chapter 3 and in [160] resulted in not being accurate enough for the selection of a core-pixel in each jth window without any previous knowledge, modifications were, therefore, required. A constraint on the selection of the possible pixel, based on the proximity to the previously chosen pixels, was described in Chapter 3, which causes a dependency on the first selected core-pixel which is undesirable. Therefore, to avoid this dependency we introduced two new features.

- *Respiration rate clusters (RR clusters)*: the second feature comes from the consideration that pixels containing the respiration thermal signature are clustered in groups with similar frequencies, i. e. the RR. Therefore, for our second feature, we select the frequency corresponding to the highest peak in the spectrum as:

$$rr_{m,l} = \arg\max_{f_k} \left(\mid y'_{m,l}(f_k) \mid \right). \tag{4.2}$$

The $rr_{m,l}$ are arranged back into the image shape, $\mathbf{RR}$ called *Frequency Map* with dimensions $M \times L$, on which a non-linear spatial filter with a 3×3 kernel is applied, as follows:

$$w_{m,l} = \frac{1}{9} \sum_{r=1}^{3} \sum_{o=1}^{3} \left(\frac{1}{exp(\kappa_1 \mid rr_{m,l} - rr_{r,o} \mid / rr_{m,l})} \right), \tag{4.3}$$

while r and o identify the kernel cell, m and l indicate the current central pixel in the entire $M \times L$ image. The constant κ_1 was set to 70, it was found to be not very critical and chosen such that it results in a weight equal to around 0.5 for a 1% relative error. Therefore, the resulting matrix will map the pixels having similar frequencies in their neighborhood and it is called $\mathbf{W}$, an example can be seen in Figure 4.4.

- *Gradient:* this feature assumes that respiration motion is visible in the thermal recordings and, considering that this is only visible if there is a

thermal contrast, an edge map is built using the gradient as follows:

$$g_{m,l} = \begin{cases} 1 & \text{if } \sqrt{\left(\frac{\partial \bar{a}_{m,l}}{\partial m}\right)^2 + \left(\frac{\partial \bar{a}_{m,l}}{\partial l}\right)^2} > \kappa_2 \\ 0 & \text{otherwise,} \end{cases} \tag{4.4}$$

with $\bar{\mathbf{A}}$ being a thermal image in degree Celsius representative of the current window j and evaluated as the average of all the thermal images in the window, $\bar{a}_{m,l}$ is an element of $\bar{\mathbf{A}}$. A gradient of at least $1°C$ is considered and, therefore, the constant κ_2 is equal to $1°C$. The gradient operation is performed with the central difference method, using the standard MATLAB function *gradient*. The resulting matrix is called $\mathbf{G}$, Figure 4.4 shows an example.

Once the features are calculated, the matrices, $\mathbf{Q}$, $\mathbf{W}$, and $\mathbf{G}$ are normalized to have values between 0 and 1 and then multiplied element-wise, obtaining a combination of the three features, called $\mathbf{V}$.

The core-pixel that is assumed to contain a respiratory thermal signature is then selected as the pixel corresponding to the maximum of $\mathbf{V}$ as:

$$(m_{p_r}, l_{p_r}) = \underset{(m,l)}{\arg\max} \left(v_{m,l}\right). \tag{4.5}$$

Once this pixel is selected, the other pixels containing respiration, both flow and motion, can be found based on the Pearson's correlation coefficient. More formally, the time-domain signals, $x_{m,l}(nT_s)$, are filtered at this point with a Butterworth bandpass filter from 30 to 100 BPM, corresponding to the normal RR range in infants including also tachypnea cases [134], this filtered signal is called $\hat{x}_{m,l}(nT_s)$. The Pearson's correlation coefficient calculation on these DC-free signals as:

$$c_{m,l} = \frac{\sum_{t=1}^{N} \left(\hat{x}_{m_{p_r},l_{p_r}}(t)\right)\left(\hat{x}_{m,l}(t)\right)}{\left(\sum_{t=1}^{N} \left(\hat{x}_{m_{p_r},l_{p_r}}(t)\right)^2 \sum_{t=1}^{N} \left(\hat{x}_{m,l}(t)\right)^2\right)^{1/2}}. \tag{4.6}$$

The $c_{m,l}$ indicate the correlation between the chosen core-pixel (m_{p_r}, l_{p_r}) and a pixel in position (m, l). $\hat{x}_{m_{p_r},l_{p_r}}(t)$ and $\hat{x}_{m,l}(t)$ are the filtered time-domain signals of the core-pixel and a pixel in position (m, l) respectively. t is an index that sweeps the time samples. In this chapter, we considered pixels with an absolute correlation higher than $\kappa_3 = 0.7$, empirically chosen:

$$\mathbf{P} = (m, l) \ : \mid c_{m,l} \mid > \kappa_3. \tag{4.7}$$

$\mathbf{P}$ consists of a set of pixels that are assumed to contain a respiratory thermal signature. Figure 4.4 shows an example of all the features.

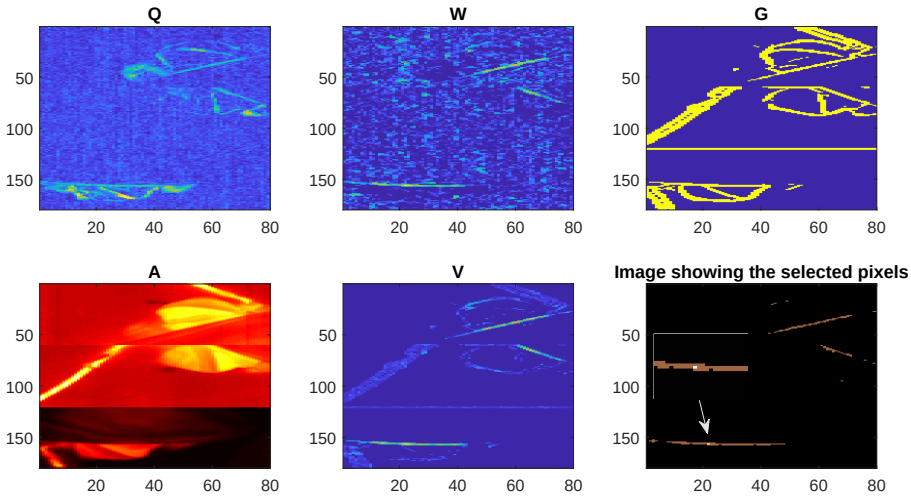


Figure 4.4: On the top row an example of the three features used, on the bottom row a representative thermal image, the result of the multiplication, and a representation of all the pixels contained in $\mathbf{P}$. The core-pixel (m_{p_r}, l_{p_r}) is shown in white in the last figure.

All pixels contained in $\mathbf{P}$ are combined together with an average operation, after correcting for the sign of each time-domain signal. The result is a single time-domain signal representing the respiration signal detected in the jth window. The signal is Hanning windowed and the spectrum is obtained through DFT. The RR is estimated as the frequency corresponding to the spectrum's peak for each window. A final time-domain signal is obtained using an overlap-add procedure as explained in [155].

Benchmarking and reference

We benchmark our multi-camera and multi-feature ROI detection algorithm against Pereira's method [102] applied on all the three different views available in our dataset. A large spatial averaging is applied in Pereira's method, which was impossible to reproduce on our images due to the very different starting resolutions (our images have a resolution of 60×80 pixels, whereas the images used in [102] have a resolution of 1024×768 pixels). Therefore, this step had to be skipped, as a consequence the quantization errors may be stronger in our dataset than in Pereira's set after down-scaling. Afterwards, a Hamming window is applied as done originally and at this point Pereira's method removes the DC component without further filtering. Directly using Pereira's method on our data resulted in SQIs being higher for the noise pixels than the pixels showing the respiration signal.

Consequently, the method failed as the wrong pixels were selected. This must be due to two main differences in our setup: the lower spatial resolution, which prohibits further downscaling, and the lower frame rate. Particularly, Pereira's method relies on spectrum features, such as the one called F2, that depend on the frame rate. Therefore, we had to optimize Pereira's method to work with our data.

We were forced to apply a filter to attenuate the low part of the spectrum in order to get meaningful results. We used a differential filter as in our method, this filter does not change the essence of Pereira's method, so we feel it is a fair adaptation. DFT is applied to obtain the spectrum, and the magnitude of each spectrum is normalized for the maximum as done in Pereira's method. The SQI needs to be calculated at this point, in [102] three frequency bands were defined, a low-pass band below 0.1 Hz, a band-pass between 0.1 Hz and 3 Hz and a high-pass band above 3 Hz, and the SQI is calculated using four empirical spectrum features (i. e. the paper [102] mentions that F1 is the maximum amplitude in the high-pass band, F2 is the percentage of values in the high-pass band that are larger than an empirical threshold, F3 is the difference between the maximum amplitudes in band-pass and low-pass bands, and F4 is the ratio between the maximum amplitudes in the low-pass and band-pass bands). Pereira et al. empirically chose a threshold, applied on the SQI, equal to 0.75 to eliminate the pixels that do not contain a respiratory signal. The three frequency bands so defined allow using the method also for the detection of heart rate (HR), whose typical frequencies are, therefore, included in the band-pass band. In our case, HR detection is not one of the application's goals and also is not practically implementable due to our limited sampling frequency which would not respect the Nyquist–Shannon theorem (i. e. the sampling frequency should be at least two times the highest frequency of a band-limited signal, for HR detection in infants upper limit up to 5 Hz can be considered [72]). Therefore, we adapted the band to the same band we use in our application (low-pass below 0.5 Hz, band-pass between 0.5 Hz and 1.7 Hz, and high-pass higher than 1.7 Hz).

The SQI has been calculated using the four features indicated in the original paper but some thresholds that were empirically defined had to be slightly adapted. Therefore, (i) the threshold for the calculation of the feature called F2 was increased from 0.1 to 0.2 and (ii) all the pixels having an SQI higher than 0.5 have been further processed for the RR calculation. We estimated the RR for each of the valid pixels and combined them using the median, obtaining an RR estimated for each window. The optimization of the SQI threshold for all the views and babies was the most problematic, as it resulted in a trade-off between the inclusion of many noise pixels, and obtaining not a number (NaN) caused by the absence of pixels meeting the criterion. We realized that using a higher threshold, e. g. 0.6 would be favorable resulting in a reduction of the error in some of the recordings, however, the use of a higher threshold could not prevent NaN. Therefore, even though by tuning the thresholds for every single view and every

single baby the results may get better, we decided to use the same thresholds for all views which, in our opinion, is more realistic.

The RR was also estimated for the simultaneously collected CI signal in the same way, the CI signal was firstly filtered with a Butterworth bandpass filter from 30 to 100 BPM. The RR so obtained constitutes our reference for both our method and Pereira's. Moving mean and median filters with 9 points were then applied in all cases.

To compare the results with the reference, root mean square error (RMSE), mean absolute error (MAE), and a percentage of correct estimates (PR) have been calculated. PR is defined as the percentage of estimates in which the absolute difference between an estimated RR and the reference RR is below 2 BPM, corresponding to the accuracy of a DFT with 4 BPM resolution. More formally:

$$PR = \frac{\sum_{j=1}^{N_w} \left(\mid rf_j - \widetilde{rf_j} \mid < 2 \right)}{N_w} \cdot 100, \tag{4.8}$$

where N_w is the number of windows, and rf_j and $\widetilde{rf_j}$ are respectively the CI's RR and the estimated RR (ours or Pereira's) obtained in the jth window. We analyzed the contribution of each of our features used for the detection of the core-pixel, and how the error varies when combining them. Moreover, to prove the dependency on the window size of features based on the spectrum shape, we analyzed how the average MAE changes by keeping all the thresholds values and changing the window size to 8 s, for our method and for the benchmark one.

4.3 Results

Table 4.2 provides the MAE, RMSE, and PR for our method and the benchmark applied on all the camera views. On the last row the average results considering all the infants is presented, the best results for MAE, RMSE, and PR are indicated in bold. Our method yields an average MAE equal to 2.07 BPM and an average PR of 70.90%, whereas the benchmark obtains its best results on camera 3 with a MAE equal to 3.18 BPM and a PR of 56.30% (we also tested the benchmark on camera 3 at full resolution, i. e. 120×160, the results were similar with a MAE equal to 3.39 BPM). Figure 4.5 shows an example of results obtained with the method proposed in this chapter and the benchmark method. A correlation and Bland–Altman analysis was also performed for our method against the CI reference and displayed in Figure 4.6. A mean bias of -0.55 BPM was obtained with the limits of agreement equal to -6.25 BPM and 5.14 BPM, and the correlation plot shows the agreement between our estimated RRs and the reference ones obtaining a $\rho = 0.97$.

Figure 4.7 shows how the individual features, used to select the core-pixel, influence the performance. The benchmark results of camera 3 are included in

Table 4.2: Results showing MAE in BPM, RMSE in BPM, and PR for our method and the benchmark one.

Infant	Ours			Benchmark								
				Cam 1			Cam 2			Cam 3		
	MAE	**RMSE**	**PR**	**MAE**	**RMSE**	**PR**	**MAE**	**RMSE**	**PR**	**MAE**	**RMSE**	**PR**
1	0.64	1.08	92.37	2.16	3.06	66.11	4.51	5.76	42.02	1.50	2.16	78.76
2	2.89	3.37	52.39	4.22	5.57	37.40	4.40	5.31	40.03	3.99	4.72	38.79
3	4.41	5.79	35.06	7.34	8.82	15.78	7.58	9.41	21.24	4.43	5.85	36.83
4	1.55	2.72	84.30	4.55	6.00	42.04	2.41	3.38	67.38	1.84	3.15	82.23
5	1.54	2.50	81.74	15.43	16.46	8.18	2.87	4.02	66.95	4.51	5.62	46.09
6	1.70	2.39	72.33	NaN	NaN	27.20	5.50	7.15	40.68	3.45	4.39	51.25
7	1.73	2.17	78.10	3.34	4.10	48.90	4.37	5.50	53.38	2.54	3.37	60.16
Avg.	**2.07**	**2.86**	**70.90**	NaN	NaN	35.09	4.52	5.79	47.38	3.18	4.18	56.30

this figure to allow comparison. The results obtained with a window of 8 s are also shown in Figure 4.7, the average MAE with this reduced window for all the babies in our method was 2.19 BPM, and 5.18 BPM for the benchmark applied on camera 3.

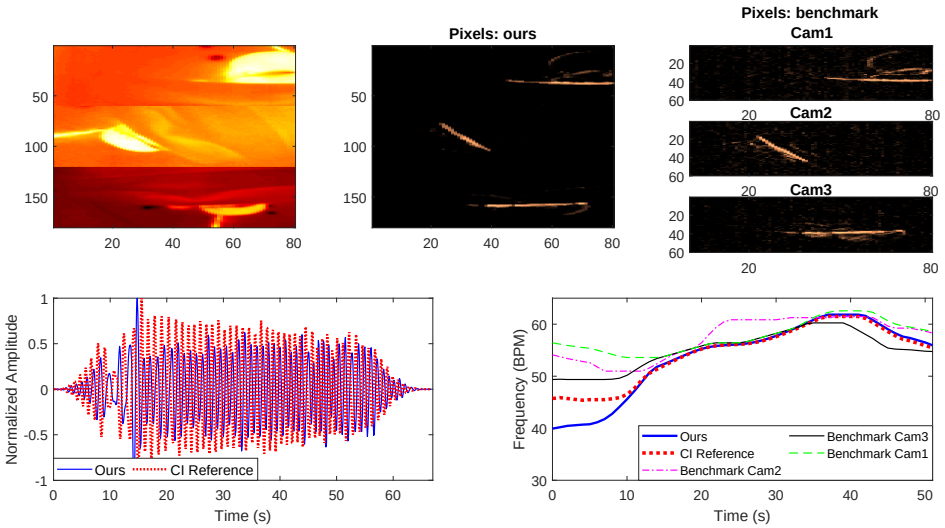


Figure 4.5: Example of results obtained on Infant 1, on the top row the first image shows the merged thermal images from the three views, the second one and the third one show respectively the pixels selected for the respiration signal estimation with our proposed method and the benchmark method. On the bottom row instead the normalized time-domain signals from the CI reference and our method and the RRs estimated in all the methods.

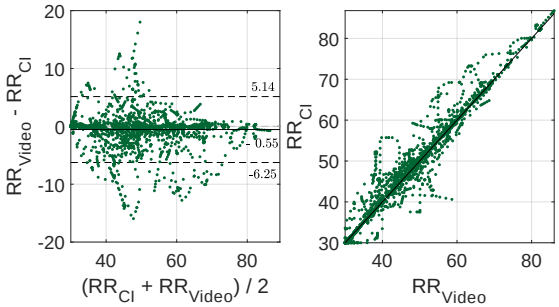


Figure 4.6: Bland–Altman plot and correlation plot. RR_{CI} and RR_{Video} are in BPM.

4.4 Discussion

In this chapter, a new algorithm has been proposed to retrieve the RR without facial landmark detection in low-resolution thermal recordings by combining pixels that highly correlate with a core-pixel that was carefully selected based on three new features. Moreover, we introduced the use of multiple parallel camera views, with data fusion on the pixel level, to enable respiration detection regardless of the momentary position of the babies. Our method was benchmarked against an adaptation of the method proposed by Pereira et al. [102].

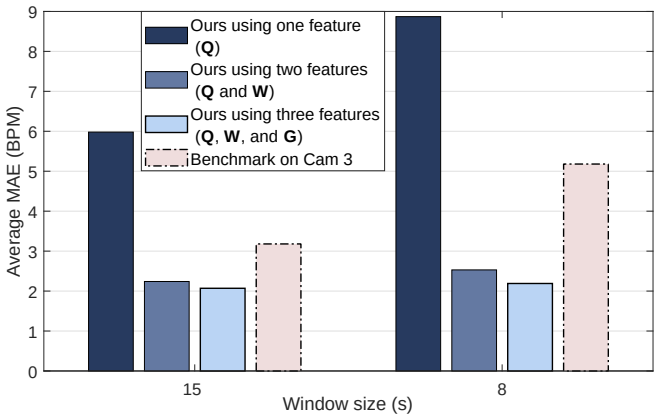


Figure 4.7: Bar plot showing the average MAE obtained when using only the *Pseudo-periodicity* feature (**Q**) or fusing it with the *RR clusters* (**W**) and the *Gradient* (**G**) features. An analysis on how the average MAE changes due to the window size in our method and the benchmark is also included.

The results obtained from around 152 minutes of infants' thermal videos proved that the method is able to correctly identify the pixels containing the respiration thermal signature, yielding an average MAE of 2.07 BPM and around 71% of correct estimates, i. e. error below 2 BPM. The highest MAE was obtained in the case of Infant 3, this infant showed a periodic breathing (PB) pattern, which is a form of immature breathing typically present in newborns [18] causing fast changes in momentary RR. Our results were better than the benchmark, regardless of the camera view selected (the benchmark performed still worse than our method, even when the combined pixel data was provided as an input, and the benchmark algorithm again optimized for this different input). Camera 3 obtained the best benchmark result, this was expected since camera 1 and 2 may contain segments in which no flow or motion are visible in one of them due to the infant position. We further emphasize that to obtain the benchmark results, Pereira's method had to be further optimized in order to get meaningful results with our data.

The Bland–Altman plot in Figure 4.6, shows the ability of our method to correctly estimate the RR, obtaining a mean bias of -0.55 BPM. The limits of agreement give an indication of the error spread $[-6.25; 5.14]$ BPM. Outliers are also present, these can be caused by synchronization problems that could be present in the data, considering that infants can have a fast changing RR small delays can cause high errors. Moreover, we should consider that the reference signal, i. e. the CI, is also prone to errors, e. g. due to loose electrodes and baby movements.

As visible from Figure 4.7, using only the *Pseudo-periodicity* feature was not reliable enough to select a core-pixel in each window, the use of a second and a third feature significantly improved the results. Moreover, our combination of three features proved to be more robust to spectrum changes due to a different window size yielding for a window of 8 s a MAE equal to 2.19 BPM compared to the significantly worsened results of the benchmark on camera 3, 5.18 BPM. We attribute this result to the use of *RR clusters* and *Gradient*, that are not dependent on the spectrum shape.

Our low-resolution cameras were able to detect both respiratory motion and flow when the relative position of the baby and the camera were favorable. Babies in open bed are a particular population in neonatal wards and their habits and sleeping patterns can significantly differ from babies in incubators. Infants in open bed are usually covered with blankets which can end up hiding the motion happening underneath. In two babies out of nine both flow and motion were not detectable, this was explained by the blanket position which ended up hiding the respiratory motion while the cameras did not have a good position for the detection of the flow. A baby with respiratory support was also included in this study, Infant 1, and resulted in having the lowest MAE compared to all other infants. this baby had a particularly quiet sleep compared to the others and the presence of the nasal cannula introduced a thermal contrast region on the face itself, increasing the number of pixels containing respiration motion.

One of the most evident limitations is that we isolated around 152 minutes of useful moments from a total amount of 19 hours corresponding to the 13.3% of the time. This can be due to several reasons, firstly, events such as feeding, diaper changing, and nursing care took up to one hour with the baby being completely out of bed. Secondly, the algorithm is currently not able to cope with significant motion of the baby or another person in the field of view. This implies that all segments including random motion of the baby, or nurses' or parents' hands in the field of view cannot be used. Since the babies included in this study were in open bed in a single-family room, where the presence of parents is highly encouraged to develop parent-infant bonding, segments where parents are comforting their baby (i. e. hands in the field of view) were quite frequent, as we shall discuss later. The occurrence of these moments could be higher compared to the case of babies in incubators. Furthermore, also, the use of a soother is quite common in this population and causes a motion pattern that can result in being similar to respiration patterns but at a different frequency, these moments have, therefore, been excluded too.

A recent work studying a vast dataset of RGB video recordings of infants in incubators [79] showed after the annotation of 384.3 hours that in 54.9% of these the reference (CI) was poor, in the 11.5% the infant was out of the incubator, and in the 11.3% there were clinical interventions, and therefore, following our classification 22.3% would remain as useful data. CI is usually disturbed during baby motion and when there is poor electrodes contact. Considering the different infant populations of the two studies we could attribute our reduced percentage of valid data to the use of the soother and to the increased parent presence. However, a more detailed annotation should be performed to analyze thoroughly the possible coverage obtainable with this type of setup and to investigate how to improve the algorithm or the setup for a more realistic clinical application, this is detailed in Chapter 5.

This pilot study gave us good insights on the use of such technologies in a complex environment as a neonatal ward. Even though the focus of this chapter was to verify the accuracy of our algorithm in detecting the RR, there is evidence suggesting this type of setup could also be used to detect the occurrence of apneas and other patterns as shown in the examples in Figure 4.8. COBs were observed in Infant 6. These were clearly visible in the video thanks to the mattress changing temperature due to the respiratory flow. Infant 3 showed a typical PB pattern throughout most of the measurements.

The detection of such events needs further analysis and should be part of future development. Moreover, our algorithm currently detects the respiratory thermal signature, without distinguishing between motion and flow, this can be a limitation when aiming at apnea detection and should be further studied, Chapter 6 proposes a solution to this problem. Our method was tested in a neonatal ward setting including only patients in an open bed, the method, the setup, or the algorithm may need adaptation to work in a NICU setting with babies in incubator, due to

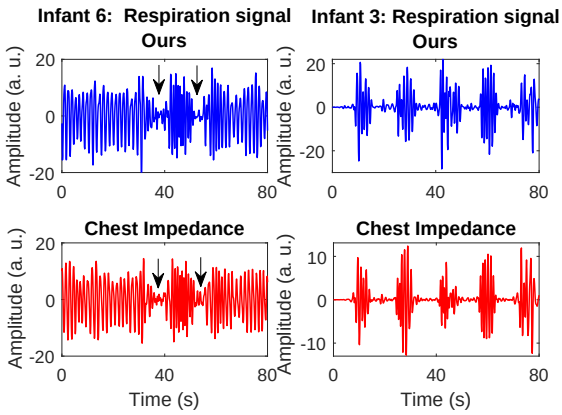


Figure 4.8: Example of detectable COBs, the figures represents two segments from Infant 6 and 3, the first row contains the respiration signal obtained with our method and the second row shows the one obtained from the CI reference. Short apneic events are indicated with the arrows and the PB pattern is clearly visible in Infant 3.

the warm and humid environment inside incubators.

4.5 Conclusions

A novel approach for automatic respiration pixels detection in multi-camera thermal recordings was introduced. Our approach is based on the merging of the three camera views and the use of three features for the detection of the pixels containing respiration motion or flow. This type of approach has the advantage of being independent of the nose visibility required in approaches based on facial landmark detection and tracking. Our method was benchmarked against the method developed by Pereira et al., adapted to work with our recordings, and both were compared to the RRs obtained with the CI, on 152 minutes recordings acquired on seven infants. Our method yielded a PR of around 71% compared to the benchmark best results of around 56%. Our algorithm obtained a MAE equal to 2.07 BPM by comparing our estimated RRs to the ones obtained from the CI reference, and using more features, not only dependent on the spectrum shape, guarantees robustness to window size changes. The algorithm proposed in this chapter will be used with adaptations in the following Chapters 5–6.

CHAPTER 5

Camera-based respiration monitoring and motion analysis

Abstract

Aiming at continuous unobtrusive respiration monitoring, motion robustness is paramount. However, some types of motion can completely hide the respiration information and the detection of these events is required to avoid incorrect rate estimations. Therefore, this chapter proposes a motion detector optimized to specifically detect severe motion of infants combined with a respiration rate detection strategy based on automatic pixels selection, which proved to be robust to motion of the infants involving head and limbs. A dataset including both thermal and red-green-blue (RGB) videos was used amounting to a total of 43 hours collected on 17 infants. The method was successfully applied to both RGB and thermal videos and compared to the chest impedance signal. The mean absolute error (MAE) in segments where some motion is present was 1.16 and 1.97 breaths per minute (BPM) higher than the MAE in the ideal moments where the infants were still for testing and validation set, respectively. Overall, the average MAE on the testing and validation set are 3.31 BPM and 5.36 BPM, using 64.00% and 69.65% of the included video segments (segments containing events such as interventions were excluded based on a manual annotation), respectively. Moreover, we highlight challenges that need to be overcome for continuous camera-based respiration monitoring. The method can be applied to different camera modalities, does not require skin visibility, and is robust to some motion of the infants.

This chapter is based on:
I. Lorato, S. Stuijk, M. Meftah, D. Kommers, P. Andriessen, C. van Pul, and G. de Haan, "Towards continuous camera-based respiration monitoring in infants," Sensors, vol. 21, no. 7, pp. 1–18:2268, 2021.

5.1 Introduction

Vital signs need to be monitored in specific hospital environments. Infants, in particular, may need continuous monitoring when admitted to neonatal wards like neonatal intensive care units (NICUs). Commonly monitored vital signs include heart rate (HR), respiration rate (RR), blood oxygen saturation, and skin temperature. Respiratory instability in infants is one of the main reasons for admission. Therefore, respiration is monitored in neonatal wards to detect critical situations like apneas, i. e. sudden cessations of breathing (COBs). If leading to hypoxia, these events can result in long-term or permanent impairment [116], and therefore, the detection of apneas is crucial.

The monitoring of respiration, but in general of most vital signals, requires attaching electrodes and sensors to the infants' skin, which can be uncomfortable for the infants or even cause skin damage [161]. Moreover, impedance pneumography or chest impedance (CI), which is commonly used in neonatal wards for respiration monitoring, is not very reliable in apnea detection [23]. For these reasons, unobtrusive solutions are being investigated for both hospital environments and home-care. Respiration motion can be detected using red-green-blue (RGB) or near-infrared (NIR) cameras [73, 76, 78], radars [57, 124, 125], or pressure-sensitive films [66, 129, 162]. Solutions using thermal cameras as in mid-wave infrared (MWIR) or long-wave infrared (LWIR) have also been investigated [97, 101, 106]. Thermal cameras can detect both respiration motion and respiratory flow, which can be useful in the detection and identification of apnea episodes in infants since obstructive apneas (OAs) and mixed apneas (MAs) still present respiratory effort, i. e. motion, but no flow [115].

Motion artifacts are a major problem for both the current monitoring technologies, e. g. CI, and most of the non-contact solutions [163, 164]. Motion robustness is, therefore, paramount when aiming at a continuous RR detection in infants. Moreover, since lethargy (hypotonia and diminished motion) and seizures (epileptic insult, repetitive motion activity) are associated with serious illnesses of the newborn [165, 166], motion is an important vital sign, that has also been linked to the prediction of apnea and neonatal sepsis [167, 168].

Multiple works proposed solutions to tackle the motion artifacts or random body movement problem in camera-based respiration detection [75, 77]. However, not all random body movements hide the respiration information and by excluding all the segments containing motion from the respiration monitoring step, potentially usable segments are also excluded. In a recent study published by Villarroel et al. [79] motion robustness was achieved by combining an indicator of the quality of the reference signal with an indicator of the agreement between the RRs obtained using different sources. However, the detection of the respiration signals is dependent on skin visibility. Infants who are cared for in open beds in neonatal wards or in home-care environments are usually covered with blankets and wear clothing. A solution based on skin visibility, particularly of the chest/torso

area, would, therefore, be impractical for these cases.

Therefore, extending Chapter 4, which estimated the RR in static moments extracted from infants' thermal videos, we analyze the performance of our algorithm in challenging conditions containing various types of motion, also semi-periodic ones such as non-nutritive-sucking (NNS). We aim at achieving motion robustness by ensuring that the RR can be accurately estimated also in the presence of some motion, e. g. head and limbs movements. We achieve this using a motion detector optimized to detect specifically the kind of motion hiding the respiratory information, which often cause impaired CI reference signal as well. This algorithm was trained and tested on thermal and RGB videos, both video types were acquired on different infants, i. e. the babies in the thermal videos are different from the babies in the RGB videos. In total, the thermal dataset includes around 42 hours of video recorded on fifteen infants in a neonatal ward. The RGB dataset is smaller and includes 50 minutes of video recorded on two infants. We, therefore, prove that both our motion detector and our RR estimation algorithm with improved motion robustness can be used for both visible and thermal modalities, without the need of skin visibility.

The remaining of this chapter is organized as follows: Section 5.2 describes the method developed and explains the setup used and the dataset. Sections 5.3 and 5.4 present the results obtained and the discussion, respectively. Section 5.5 contains the conclusions.

5.2 Materials and methods

In this section the datasets used, the manual annotation performed, and the experimental setups are detailed. Furthermore, the method for motion detection and respiration monitoring are described.

5.2.1 Materials

Both thermal and RGB datasets used in this chapter were collected in the neonatal ward of the Máxima Medical Centre (MMC) in Veldhoven, The Netherlands. Both studies received a waiver from the ethical committee of the MMC (the thermal dataset with ID N19.074 and the RGB dataset with ID N12.072), and informed consent was obtained from the infants' parents prior to the study.

Dataset

The dataset was split into two sets, one called the training and testing set, which is used to optimize and test the motion detection step, and adjust our respiration monitoring algorithm. The other one called the validation set will be used to obtain unbiased results for both the motion detection step and the RR detection

Table 5.1: Infants' data for the training and testing set (indicated with T&T) and the validation set (indicated with V). The gestational age (GA) is indicated in weeks (w) and days (d), postnatal age (PA) is in days.

Infant	Video Type	GA (w+d)	PA (days)	Sleeping Position	Duration (hours)	Set
1	Thermal	26w 4d	59	Supine	2.98	T&T
2	Thermal	38w 5d	3	Supine	2.74	T&T
3	Thermal	34w 1d	16	Supine	2.93	T&T
4	Thermal	26w 3d	59	Prone	3.16	T&T
5	Thermal	39w	2	Lateral	3.05	T&T
6	Thermal	40w 1d	6	Supine	2.95	T&T
7	Thermal	40w 2d	1	Lateral	0.92	T&T
8	RGB	36w	47	Supine	0.30	T&T
9	RGB	30w	34	Supine and Lateral	0.57	T&T
10	Thermal	26w 4d	77	Supine	2.94	V
11	Thermal	26w 4d	77	Supine	2.97	V
12	Thermal	33w 4d	5	Supine	2.97	V
13	Thermal	34w 2d	9	Supine	2.87	V
14	Thermal	32w 2d	11	Supine	2.96	V
15	Thermal	35w 1d	8	Supine	2.94	V
16	Thermal	38w 1d	2	Supine	3.00	V
17	Thermal	27w 5d	16	Supine	2.96	V

step. Table 5.1 contains the infants' data and the duration of the recordings for the training and testing set, and the validation set. The infants were assigned to the two sets based on the availability of the data. The thermal videos, partially already used in Chapter 4, amount to a total of around 42 hours acquired on fifteen infants, all the infants were monitored for around 3 hours except for infant 7, which has a total video duration of around 1 hour, due to setup problems. The RGB videos of infants 8 and 9 amount to a total video duration of around 52 minutes.

Manual annotation

The videos content was annotated, including the occurrences of motion events, and it was then used as ground truth for the motion detection step. A MATLAB built-in application called *Video Labeler* was used to annotate the videos. A set of labels was defined to describe the possible visible events, the labels are not exclusive, meaning that multiple labels can be true at the same time. We defined two classes of motion type 1 and type 2 motion. The labels are presented in Table 5.2.

The main difference between the two types of motion, i. e. 1 and 2, is the involvement of the chest in the motion event. Type 1 is a motion that involves the chest/torso area, where the respiration motion can be usually seen. In our

Table 5.2: Labels used for the manual annotation.

	Annotation Labels	Subcategories and Details
Included	**(i) Infant activity**	• Still • Type 1 motion (motion including chest/torso area) • Type 2 motion (motion involving limbs or head)
	(ii) NNS	-
Excluded	**(iii) Interventions**	includes both parents' and caregivers' interventions
	(iv) Other	• Someone in the background • Baby out of bed • Camera motion • Unsuitable view

classification, this is, therefore, considered as the kind of motion that results in hiding the respiration information, which can cause artifacts also in the CI reference signal. Type 2, instead, does not involve chest or torso movements but affects other parts as, head, hands, arms, fingers, or even facial expressions.

The segments of videos including events labeled as categories iii and iv in Table 5.2 were excluded in this work since they would require different detectors, e. g. interventions detection or infant presence detection [79]. In particular, the included and excluded percentages in the entire dataset are, respectively, 73.86% and 26.14%. The majority of the excluded moments are caused by the babies being out of bed and by interventions, 46.4% and 31.8%, respectively. The breakdown of the included moments is shown in Figure 5.1 split between the training and testing, and the validation set.

The segments containing type 1 motion events are considered unusable for the estimation of the RR, whereas the ones containing type 2 motion, still, and NNS are considered usable. The cumulative percentages of type 2 motion, still, and NNS constitute 70.03% and 68.85% of the included moments for the training and testing, and the validation set, respectively. The remaining part contains the fragments annotated with type 1 motion. The occurrence of type 1 motion is, therefore, very similar between the two sets.

Experimental setup

Two different setups were used to collect the RGB videos and thermal videos used in this chapter. The thermal videos were collected using three thermal cameras positioned around the infants' bed. The cameras used are FLIR Lepton 2.5, they are sensitive in the LWIR range, the resolution is 60×80 pixels, the thermal

sensitivity is 50 mK, and the average frame rate is 8.7 Hz. The acquisition was performed using MATLAB (MATLAB 2018b, The MathWorks Inc., Natick, MA, USA). Due to the acquisition strategy, the 3 hours of recording are split into 9 videos of 20 minutes each, gaps of up to 4 s can be present between the videos. For further information on the setup refer to Chapter 4.

The visible images were obtained in a separate data collection with a single RGB camera (UI-2220SE, IDS), that was positioned on a tripod to visualize the baby in the open bed. Some videos were collected from the side and others from the top. The frame rate and resolution are, respectively, 20 Hz and 576 × 768 pixels. In both cases, the reference RGB signal sampled at 62.5 Hz was collected using the patient monitor (Philips MX800). To solve the synchronization problem, an artifact (simultaneously disconnecting the CI leads and covering the view of one of the cameras) was generated at the start of each recording to synchronize CI and videos.

5.2.2 Methods

The algorithm proposed in this chapter can be split into two main parts, i. e. motion detection and RR estimation. The first was designed to detect type 1 motion, since segments containing type 2 motion are considered usable for the RR detection and it is, therefore, not necessary to detect their occurrence. Therefore, if type 1 motion was detected the RR could not be accurately estimated and an indication that the baby was moving was provided. Otherwise, the video segment did not contain type 1 motion and it was classified as usable and the RR was estimated using the second part of our algorithm. These steps are shown in Figure 5.2. The algorithm was implemented using MATLAB.

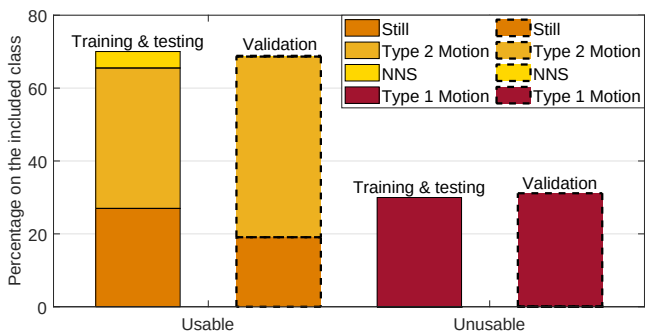


Figure 5.1: Results of the manual annotation: the breakdown of the included class into the subcategories for the training and testing, and the validation set.

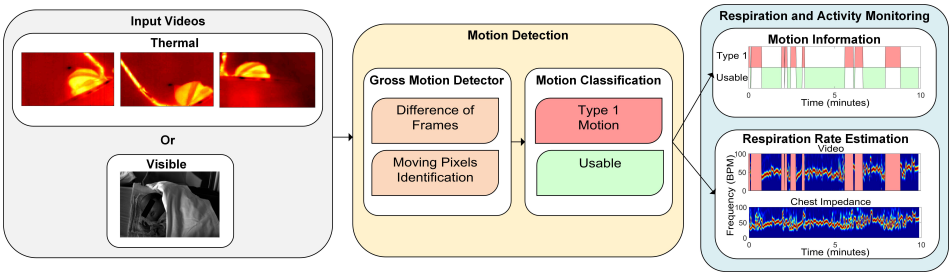


Figure 5.2: Main blocks of the processing chain and an example of the results.

Preprocessing

The thermal videos were linearly interpolated to compensate for the acquisition strategy, which resulted in a non-uniform sampling rate because external triggering was not used. A 1D interpolation was applied to each pixel's time-domain signal, using the MATLAB function *interp1*, the result was three videos sampled at 9 Hz, close to the average frame rate, with a resolution of 60×80 pixels. The RGB data were converted to grayscale (using the MATLAB function *rgb2gray*) and downscaled, to allow faster processing, with a downscale factor of 3 resulting in a final video resolution of 192×256. The grayscale videos were also temporally downsampled to reach the same sampling rate as the thermal videos, i. e. 9 Hz, from an initial sampling frequency of 20 Hz, for faster processing. The frame sizes will be indicated as $\tilde{M} \times \tilde{L}$, which will correspond to 60×80 in the thermal case and 192×256 in the visible case.

A sliding window approach was used for both the motion detection and the RR estimation steps. Considering a trade-off between latency and frequency resolution and the fact that the use of longer windows means more sliding windows may contain motion events, a relatively short window size of 8 s was chosen with a slide of 1 s. A simplified diagram of the motion detection and the respiratory pixels identification algorithms is shown in Figure 5.3.

Motion detection

- Gross motion detector: let $\mathbf{X}(nT_s)$ be the frames in the jth window, with $n = 0 + (j-1)/T_s$, $1 + (j-1)/T_s$, ..., $(N-1) + (j-1)/T_s$, and $N = 72$ samples, corresponding to the samples in the jth window with a sampling period $T_s = 0.111$ s. The gross motion detector was based on the absolute value of the difference of frames (DOF) in the jth window. More formally:

$$\mathbf{D}(uT_s) = |\frac{\partial \mathbf{X}(nT_s)}{\partial n}|, \tag{5.1}$$

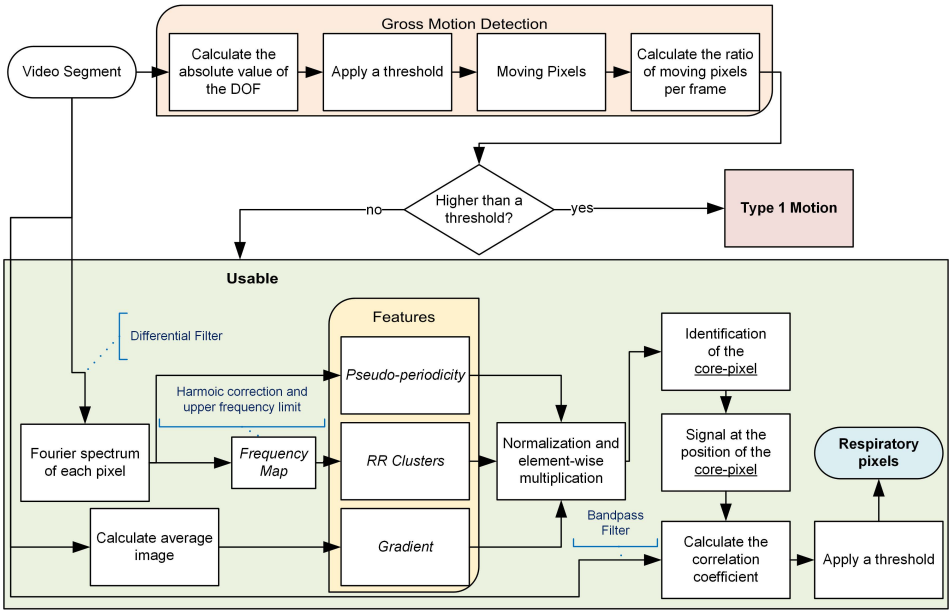


Figure 5.3: Simplified block diagram of the motion detection and the respiratory pixels selection algorithms.

the $\frac{\partial}{\partial n}$ operator represents the partial derivative with respect to the time dimension. $\mathbf{D}(uT_s)$ contains the frames resulting from the absolute value of the DOF operation at each time sample, with $u = 0 + (j-1)/T_s$, $1 + (j-1)/T_s$, ..., $(N-2) + (j-1)/T_s$. At this point, a first threshold value was introduced, which turns $\mathbf{D}$ into binary images identifying what we considered to be moving pixels:

$$\mathbf{MP}(uT_s) = \begin{cases} 1 & \text{if } \mathbf{D}(uT_s) > thr_1 \\ 0 & \text{otherwise.} \end{cases} \tag{5.2}$$

thr_1 is a threshold that was introduced to differentiate the source of the change between noise and motion, it is defined as:

$$thr_1 = \frac{Range(\mathbf{X})}{f_1}, \tag{5.3}$$

the numerator represents the range of $\mathbf{X}$, i. e. the difference between the maximum value and the minimum value considering all the pixels of all the frames in $\mathbf{X}$, and f_1 is a value that was optimized. The ratio of moving pixels was then calculated as:

$$s(uT_s) = \frac{\sum_{\tilde{m}=1}^{\tilde{M}} \sum_{\tilde{l}=1}^{\tilde{L}} mp_{\tilde{m},\tilde{l}}(uT_s)}{\tilde{M} \cdot \tilde{L}}. \tag{5.4}$$

Here, $mp_{\tilde{m},\tilde{l}}(uT_s)$ is an element of $\mathbf{MP}(uT_s)$ at the position $(\tilde{m}, \tilde{l})$.

- Motion classification: the ratio of moving pixels $s(uT_s)$ was used to perform the classification between usable and unusable segments for RR detection. In particular, we aim at detecting the unusable moments, i. e. the ones containing type 1 motion. The main assumption is that type 1 is part of a more complex kind of motion, typical of infants' crying motion. Therefore, the simplest way to detect it is to assume that type 1 motion will result in more moving pixels compared to any of the usable segments.

  To perform a classification between the two, a second threshold thr_2 was introduced, which was applied to the ratio of moving pixels $s(uT_s)$. The final classification was, therefore, performed on a window-based fashion, i. e. each window was classified as containing type 1 motion, corresponding to 1, or usable, corresponding to 0.

  Since we used three cameras in the thermal setup, we applied this algorithm three times. For the RGB dataset this was not necessary, as there was only a single camera used. In the visible case the classification will be:

$$Motion_{vis}(j) = \begin{cases} 1 & \text{if } \exists \ u : s(uT_s) \geq thr_2 \\ 0 & \text{otherwise.} \end{cases} \tag{5.5}$$

  For the thermal case instead:

$$Motion_{th}(j) = \begin{cases} 1 & \text{if } \exists \ u : (s_1(uT_s) \geq thr_2 \text{ OR} \\ & \qquad s_2(uT_s) \geq thr_2 \text{ OR} \\ & \qquad s_3(uT_s) \geq thr_2) \\ 0 & \text{otherwise.} \end{cases} \tag{5.6}$$

  $s_1(uT_s)$, $s_2(uT_s)$, and $s_3(uT_s)$ are the ratios of moving pixels obtained from the three thermal views.

- Ground truth: The ground truth used to evaluate the performance of our motion detector was obtained based on the manual annotations presented in Section 5.2.1. In particular, the ground truth was built using the sliding window approach. Each window was classified as excluded, as type 1 motion, or as usable. The condition used was the presence of at least a frame in the window which results in being true for one of those categories. The excluded class had the priority, if this was true for at least a frame in the window, the entire window was classified as excluded. If the latter was false then type 1 motion was taken into consideration in the same manner, and lastly if the two above were both false we classified the window as usable.

- Parameters optimization: the factor f_1, for the moving pixels detection, and the threshold thr_2, for the motion classification, were optimized. A

leave-one-subject-out cross-validation was used to optimize the two parameters. The approach was chosen considering that environment changes, e. g. environment temperature, blankets type, and position, can influence the parameters values and therefore, the between-baby variability is more important than the within-baby variability. The set of parameters that resulted in the highest balanced accuracy for each fold was considered as a candidate set. The final chosen set was the most selected candidate set. This metric was preferred compared to the classic accuracy due to the imbalance in our two classes (usable was more frequent than type 1 motion). The optimization was performed on the training and testing set, presented in Table 5.1. This set includes 9 babies and therefore, 9 folds were performed in the cross-validation. Two sets of parameters were empirically chosen for the training and correspond to $f_1 = [4; 5; 6; 7; 8; 9; 10; 11; 12]$ and $thr_2 = [0.004; 0.005; 0.006; 0.007; 0.008; 0.009; 0.010; 0.011; 0.012]$. The most chosen set, used in the next steps, was $f_1 = 8$ and $thr_2 = 0.005$, more information on the results can be found in Section 5.3.

Respiration rate estimation

Respiratory signal and rate were both estimated in the windows in which the motion detection step results in the usable category using an adjustment of our previous method [169], introduced in Chapter 4. Briefly, first the images of the thermal videos were merged together in a single image plane, resulting in a single video with resolution 180×80, whereas the grayscale videos were processed with the single view available, i. e. videos with resolution 192×256. These two possible frame dimensions will be referred to as $M \times L$. Our method is based on the automatic detection of the pixels containing respiration information. This is performed using the three features presented in Chapter 4 and in [169], improvements were applied to tackle new challenges highlighted by the extension of our dataset and of the acceptable motion.

The changes involve an adaptation of the second feature, *Respiration rate clusters*, adapted to overcome the presence of the respiration's first harmonic and NNS pattern in part of the extended dataset. The third feature (*Gradient*) was also adapted for the use on visible images, now added to the dataset, and finally, the correlation value that indicates which pixels contain the respiration information was increased. More in detail, each pixels' time-domain signal is indicated as $x_{m,l}(nT_s)$, with (m, l) indicating the pixel. Three features were used to find a core-pixel, in each $\hat{j}$th window, which was then employed to find (using a correlation metric) all the helpful pixels that can be combined to compute the respiratory signal, with $\hat{j} = j \ : \ Motion(j) = 0$.

- *Pseudo-periodicity*: this first feature is based on the assumption that respiration can be considered a periodic signal. This feature was not changed compared to [169]. A differential filter was used to attenuate low-frequencies

resulting in filtered time-domain signals called $x'_{m,l}(nT_s)$. The signals were zeropadded, reaching a length equal to $N_z = 120 \cdot N$, and multiplied for a Hanning window. Afterwards, a 1D discrete Fourier transform (DFT) was used to estimate the spectrum called $y'_{m,l}(f_k)$ with $k = 0, 1, ..., N_z/2 - 1$ and $f_k = k/(N_z T_s)$ Hz. This feature consists in the calculation of the height of the normalized spectrum's peak. More formally:

$$q_{m,l} = \frac{\max\limits_{0 \leq f_k \leq \frac{(N_z/2-1)}{N_z T_s}} (\mid y'_{m,l}(f_k) \mid)}{\sqrt{\sum\limits_{f_k=0}^{\frac{(N_z/2-1)}{N_z T_s}} \mid y'_{m,l}(f_k) \mid^2}}. \tag{5.7}$$

Each $q_{m,l}$ represents the height of the peak of the normalized spectrum of the pixel in position (m, l), $q_{m,l}$ are elements of the first feature $\mathbf{Q}$.

This feature is sensitive to the presence of type 2 motion. Regions moving due to this type of motion can generate a big variation in the pixels' values (depending on the contrast). This variation can, therefore, produce a strong DC component, which will result in a high $q_{m,l}$. The combination with the other features allows us to obtain motion robustness, Figure 5.4 presents an example during a type 2 motion and the *Pseudo-periodicity* feature is visible in Figure 5.4b.

- *Respiration rate clusters (RR clusters)*: this feature is based on the observation that respiration pixels are not isolated but grouped in clusters. To automatically identify the pixels of interest more accurately, modifications were introduced to this feature to improve the robustness to the presence of NNS, typical when the infant has the soother, and to cope with the presence of the respiration's first harmonic. The frequencies corresponding to the local maxima of the spectrum $y'_{m,l}(f_k)$ were found and the properties of the harmonic were checked:

$$\mathbf{h}_{m,l} = \arg\underset{lim_1 < f_k < lim_2}{\text{localmax}}(\mid y'_{m,l}(f_k) \mid), \tag{5.8}$$

$\mathbf{h}_{m,l}$ is a vector, obtained for the pixel in position (m, l), containing the frequencies of the local maxima in the band of interest, which is identified by lim_1 and lim_2 respectively 0.5 and 1.83 Hz. The length of the vector is, therefore, variable and dependent on the spectrum content of each pixel (m, l), this operation was performed using the MATLAB function *findpeaks*.

The harmonic properties were checked:

$$rr_{m,l} = \begin{cases} h_{m,l}(1) & \text{if } \exists\ \hat{z} > 1 : \\ & \mid h_{m,l}(\hat{z}) - 2 \cdot h_{m,l}(1) \mid < \frac{1}{NT_s}\ AND \\ & (y_{m,l}(h_{m,l}(\hat{z})) < y_{m,l}(h_{m,l}(1))\ AND \\ & y'_{m,l}(h_{m,l}(\hat{z})) \geq y'_{m,l}(h_{m,l}(1))) \\ \arg\max\limits_{f_k} \left(\mid y'_{m,l}(f_k) \mid \right) & \text{otherwise,} \end{cases} \tag{5.9}$$

$y_{m,l}(f_k)$ is the spectrum of the pixels' time-domain signal calculated as $y'_{m,l}(f_k)$ but without applying the differential filter and $h_{m,l}$ is an element of $\mathbf{h}_{m,l}$.

We have, therefore, estimated the main frequency component for each pixel. To avoid erroneous RR estimation caused by higher frequencies components, e. g. caused by NNS, the $rr_{m,l}$ that were higher than lim_2 were put to zero. Therefore:

$$\hat{rr}_{m,l} = \begin{cases} rr_{m,l} & \text{if } rr_{m,l} < lim_2 \\ 0 & \text{otherwise.} \end{cases} \tag{5.10}$$

The $\hat{rr}_{m,l}$ are elements of $\hat{\mathbf{RR}}$, *Frequency Map*, an example is shown in Figure 5.4f. The non-linear filter introduced in [169] was applied:

$$w_{m,l} = \frac{1}{9} \sum_{r=1}^{3} \sum_{o=1}^{3} \left(\frac{1}{exp(\kappa_1 \mid \hat{rr}_{m,l} - \hat{rr}_{r,o} \mid / \hat{rr}_{m,l})} \right), \tag{5.11}$$

where r and o identify the kernel cell, whereas m and l indicate the pixel. κ_1 is a constant empirically chosen and equal to 70 as indicated in our previous work [169] and in Chapter 4. The resulting frame $\mathbf{W}$ will map the pixels having similar frequencies around them.

It should be noted that the $\hat{rr}_{m,l}$ on which we imposed the value 0 in Equation 5.10, will not result in a high $w_{m,l}$, even if there are clusters of zeros in $\hat{\mathbf{RR}}$. This is due to the equation of the filter that with $\hat{rr}_{m,l} = 0$ will produce not a number (NaN). The same will happen for regions with type 2 motion, where the main frequency component is the DC. This property allowed to avoid type 2 motion regions in the pixel selection phase achieving motion robustness, an example is visible in Figure 5.4e.

- *Gradient*: this last feature is based on the assumption that respiration motion can be only visualized at the edges. This feature has been modified to make it independent of the setup used:

$$g_{m,l} = \begin{cases} 1 & \text{if } \sqrt{\left(\frac{\partial \bar{a}_{m,l}}{\partial m}\right)^2 + \left(\frac{\partial \bar{a}_{m,l}}{\partial l}\right)^2} > \kappa_2, \\ 0 & \text{otherwise,} \end{cases} \tag{5.12}$$

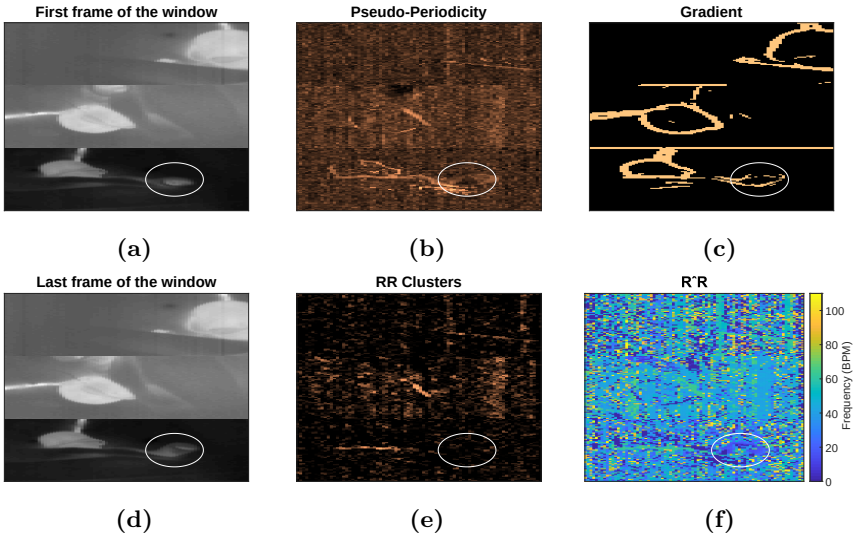


Figure 5.4: Example of features obtained during a type 2 motion, i. e. arm motion. In (**a**,**d**) the merged thermal images are presented, the circle indicates the position of the baby's arm where the type 2 motion is happening. The images in (**b**,**c**,**e**) show the three features. While in this case, *Pseudo-periodicity* and *Gradient* are sensitive to the presence of type 2 motion, *RR clusters* is not, this is due to the $\hat{\mathbf{R}}\mathbf{R}$ matrix shown in (**f**) where the arm area can have frequencies equal to zero.

where $\frac{\partial}{\partial m}$ and $\frac{\partial}{\partial l}$ represent the partial derivatives in the two spatial dimensions, and $\kappa_2 = Range(\mathbf{A})/16$ was empirically chosen since resulted in identifying the edges of both thermal and grayscale images, with $\mathbf{A}$ being the series of frames in the $\hat{j}$th window. $\bar{\mathbf{A}}$ is an average image representative of the current window $\hat{j}$ evaluated as the average of all the images in $\mathbf{A}$, with elements $\bar{a}_{m,l}$. The resulting matrix will be the third feature $\mathbf{G}$. The use of $\bar{\mathbf{A}}$ to evaluate the gradient can also ensure robustness to some type 2 motion, whose regions will not be visible in the average image if the motion is transient enough. In the example in Figure 5.4c the pixels involved in the type 2 motion are still selected in the *Gradient* feature, but *RR clusters* ensures the correct pixels are chosen.

The features, $\mathbf{Q}$, $\mathbf{W}$, and $\mathbf{G}$, were then combined together, after being normalized between 0 and 1, by multiplying them and obtaining $\mathbf{V}$. This feature combination was used to identify the core-pixel as:

$$(m_{p_r}, l_{p_r}) = \underset{(m,l)}{\arg\max} \left(v_{m,l}\right), \tag{5.13}$$

where $v_{m,l}$ identifies an element of $\mathbf{V}$. The pixels containing respiration information were then found from this core-pixel based on the Pearson's correlation coefficient, estimated using a bandpass filtered version of the pixels' time-domain signal. The filter used is a Butterworth bandpass between lim_1 and lim_2. In Chapter 4 pixels having a correlation higher than 0.7 with the core-pixel were considered to contain respiration information, this threshold on the correlation value has been increased in the current version of the algorithm considering the reduction in window size and the fact that the accuracy of the correlation estimation depends on the length of the signal. Therefore, the threshold has been set to 0.9 and indicated with κ_3. In particular:

$$\mathbf{P} = (m, l) \ : \mid c_{m,l} \mid > \kappa_3, \tag{5.14}$$

where $c_{m,l}$ is the correlation between the core-pixel (m_{p_r}, l_{p_r}) and the signal of the pixel in position (m, l), calculated using the MATLAB function *corrcoef*. $\mathbf{P}$ will, therefore, be a set indicating the pixels containing the respiration signal and can have variable dimension depending on the window $\hat{j}$. To calculate the RR and the respiration signal, all the bandpass filtered time-domain signals of the pixels in $\mathbf{P}$ were combined using an average operation. The RR was calculated from the spectrum of this signal after using a Hanning window, and the RR was estimated as the frequency corresponding to the spectrum's peak for each window. The same was applied to the CI signal to estimate the reference RR from the waveform. These spectra were then arranged into a short time Fourier transform (STFT).

Evaluation metrics

Accuracy (ACC), balanced accuracy (BalACC), sensitivity (SE), and specificity (SP) were calculated for the test step of the cross-validation and for the validation dataset to obtain unbiased performance results. The RR was compared to the one obtained using the CI. Mean absolute error (MAE), root mean square error (RMSE), and percentage of correct estimates (PR) [169], considering an accuracy of 3.75 breaths per minute (BPM) caused by the window size, were calculated. We estimated the percentage of time used (PT) by calculating the percentage of windows classified as usable by the motion classification step on the number of windows in the included data (which includes also type 1 motion occurrences).

To prove the improved motion robustness of our algorithm, we used the annotations to identify the moments containing only type 2 motion and compared them with the ones containing only stillness. Moreover, the contribution of the NNS segments to the error was also analyzed. The average MAE was obtained in all these windows to analyze their contribution to the final error. In these cases, PT is calculated by considering also the information of the manual annotation on the occurrences of specific events. For example, PT for the segments containing only type 2 motion is calculated considering the number of windows classified as usable by our motion detection and that according to the manual annotation contain only

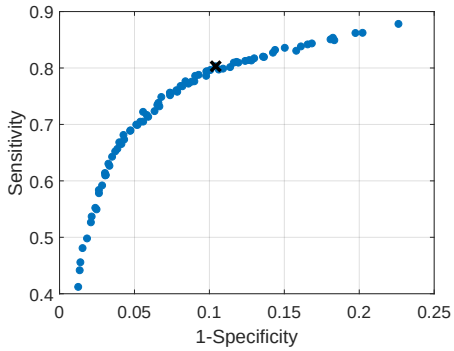


Figure 5.5: ROC curve obtained with the nine folds of the cross-validation by using all the parameters combinations.

type 2 motion, or PT in the usable segments excluding NNS is evaluated using the number of windows classified as usable and that do not contain NNS according to the manual annotation.

5.3 Results

The average receiver operating characteristics (ROC) curve for all nine folds obtained from the cross-validation applied on the training and testing set, is presented in Figure 5.5. The blue points represent the average sensitivity and specificity on all folds for that particular combination of f_1 and thr_2, whereas the cross is the average sensitivity and specificity on all folds corresponding to the most chosen parameter set. Table 5.3 shows the results of ACC, BalACC, SE, and SP using the final chosen set of parameters for the testing stage of the cross-validation and for the validation set that was not involved in the training.

The results obtained in the RR detection step are shown in Tables 5.4 and 5.5. The first one shows the MAE obtained in all moments considered usable by our motion detection step (that includes segments containing NNS) and the error in

Table 5.3: Average performance of the motion detection step for all the babies of the training and testing, and the validation set using the chosen parameters.

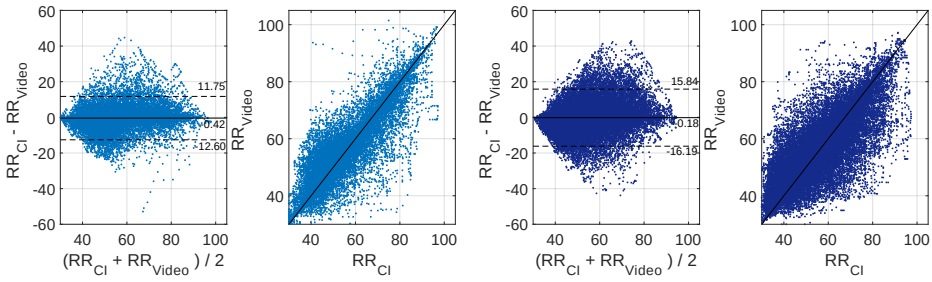
	ACC	**BalACC**	**SE**	**SP**
Training and testing set	88.22%	84.94%	80.30%	89.58%
Validation Set	82.52%	77.89%	66.85%	88.93%

Table 5.4: Average and standard deviation of MAE and PT on all babies of the training and testing set for the previous version of method [169] (Chapter 4) and the current one presented in this chapter, in the windows classified by the motion detector as usable. We further show the contribution of the NNS to the overall error (these segments were obtained using the manual annotation).

	Previous Version of Method [169]		**Current Version of the Method**	
	Usable	**NNS Only**	**Usable**	**NNS Only**
MAE (BPM)	4.54 ± 1.82	9.39 ± 3.68	3.55 ± 1.63	7.11 ± 4.15
PT	68.59% ± 13.29%	4.59% ± 6.93%	68.59% ± 13.29%	4.59%± 6.93%

the moments containing only NNS, whose windows were determined using the manual annotation. Moreover, a comparison between the respiration detection method introduced in Chapter 4 and in [169], and the modified one introduced in this chapter is also presented. Table 5.5 contains the results obtained in all the usable segments excluding the NNS windows on the two sets. Moreover, using the manual annotation, we also show the errors in the windows containing only type 2 motion, and in the ideal moments in which the infants are still.

Figure 5.6 presents Bland–Altman and correlation plots for the training and testing set, and the validation set, respectively, using the RRs in all the usable windows excluding the NNS. The mean bias were -0.42 and -0.18 BPM and the correlation plot shows the agreements between the reference and our estimation with a $\rho = 0.90$ and $\rho = 0.80$ for the training and testing set, and the validation set. Example results are presented in Figures 5.7–5.9.



(a) **(b)**

Figure 5.6: Bland–Altman and correlation plot: (**a**) training and testing set, (**b**) validation set. RR_{CI} and RR_{Video} are in BPM.

Table 5.5: Results of the two sets in the segments classified as usable by our motion detector excluding the NNS windows, obtained thanks to the manual annotation. The errors in the windows containing type 2 motion and moments where the infants were still are also included. MAE and RMSE are in BPM.

	Infant	Usable Excluding NNS				Type 2 motion Only		Still Only	
		MAE	RMSE	PR	PT	MAE	PT	MAE	PT
Training and testing	1	1.86	3.34	83.61%	70.38%	1.57	27.92%	1.51	34.61%
	2	2.87	3.97	73.71%	40.60%	2.56	20.90%	2.64	13.02%
	3	6.30	8.09	39.44%	67.83%	6.32	39.23%	6.28	24.38%
	4	4.43	6.21	60.16%	72.75%	4.99	44.09%	2.49	20.39%
	5	5.04	7.61	56.44%	40.22%	4.84	29.24%	2.24	5.35%
	6	2.97	4.73	71.34%	66.74%	3.70	29.96%	1.94	31.69%
	7	2.80	4.15	72.08%	46.16%	2.57	30.28%	0.70	4.61%
	8	1.89	3.40	88.63%	89.71%	1.76	11.47%	1.91	77.84%
	9	1.62	2.70	85.55%	81.60%	2.88	24.16%	1.08	56.76%
	Avg.	**3.31**	**4.91**	**70.11%**	**64.00%**	**3.47**	**28.58%**	**2.31**	**29.85%**
	± s.d.	**± 1.61**	**± 1.94**	**± 15.84%**	**± 17.82%**	**± 1.62**	**± 9.56%**	**± 1.62**	**± 24.22%**
Validation	**10**	4.46	6.62	61.41%	63.62%	5.52	34.40%	2.44	22.78%
	11	3.79	5.54	64.96%	55.55%	4.01	34.62%	2.27	12.29%
	12	6.23	7.98	38.98%	68.20%	5.98	33.70%	6.60	23.35%
	13	6.29	8.51	44.00%	69.53%	6.30	51.04%	3.59	6.13%
	14	6.89	9.56	47.37%	73.38%	7.35	44.73%	4.58	18.00%
	15	4.75	6.65	54.11%	78.86%	4.83	42.08%	4.39	26.81%
	16	4.09	5.73	60.97%	76.84%	4.39	28.92%	3.21	30.73%
	17	6.40	8.78	47.79%	71.22%	7.64	40.14%	3.15	19.60%
	Avg.	**5.36**	**7.42**	**52.45%**	**69.65 %**	**5.75**	**38.71%**	**3.78**	**19.96%**
	± s.d.	**± 1.21**	**± 1.49**	**± 9.35%**	**± 7.47%**	**± 1.32**	**± 7.14%**	**± 1.40**	**± 7.90%**

5.4 Discussion

Our method for motion robust respiration detection can be used for both thermal and visible modalities, and it does not rely on skin visibility or facial landmark detection. Moreover, it is able to detect motion events that are problematic for

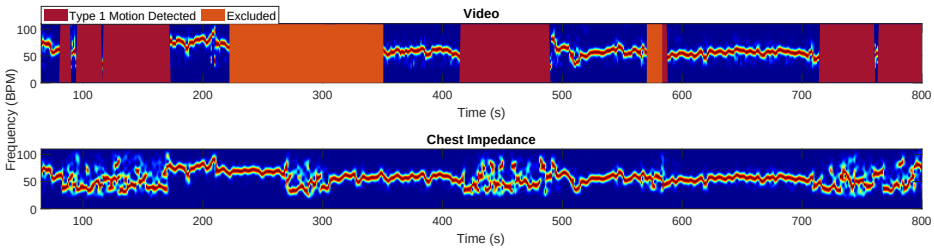


Figure 5.7: Example of the STFT obtained using the camera and the CI reference. The noisiness of the reference's spectrum during type 1 motion shows the sensitivity of the reference to this type of artifact. The excluded segments are due to camera motion.

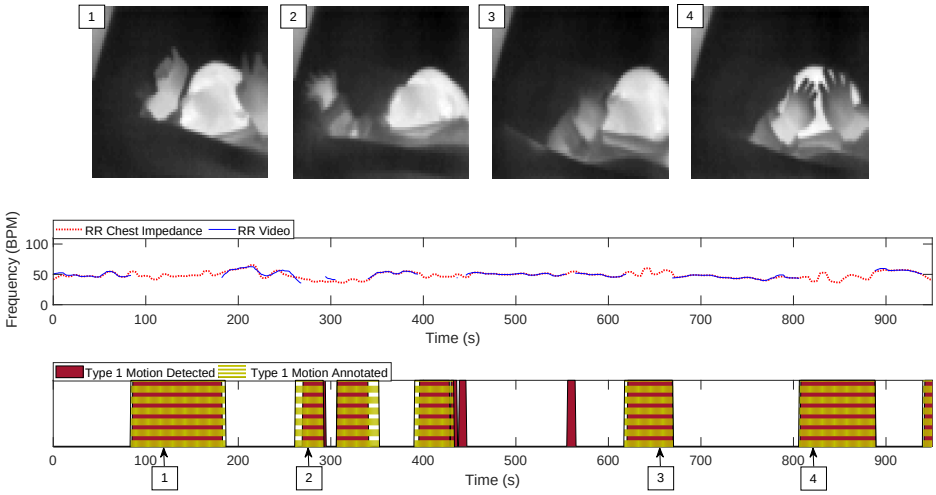


Figure 5.8: Example of results showing the RR estimated using our cameras and algorithm, and the reference one. The difference in the manual annotation of type 1 motion and the detected one is visible in the bottom plot. Examples of frames during the type 1 motion (infant crying) are also shown.

respiration monitoring, ensuring a more accurate RR detection and delivering motion information. The manual annotation showed that the RR can be potentially estimated in around 70% of the included data since the remaining 30% is annotated as type 1 motion. The impossibility to accurately estimate an RR in these segments is a limitation present in all unobtrusive technologies but also in the current monitoring modalities, i. e. CI. An example of the RRs estimated using both camera and CI in the moments annotated and automatically classified as type 1 motion is provided in Figure 5.7. The sudden noisiness in the spectrum clearly indicates the inaccuracy of the RRs estimation in these segments.

Table 5.5 shows an average PT of around 64% and 70% for the two sets, however, there is considerable variability in the PT between the infants, especially in the training and testing set, as shown by the standard deviation. Infants that are more agitated will have an increased occurrence of type 1 motion, reducing the amount of time usable for RR detection, which can be also lower than 50% (can be partially due to NNS occurrence as well). However, considering that CI is also unable to provide an RR in these cases, the information that the infant is agitated and moving frequently may be much more informative than an inaccurate estimation of RR. In addition, a patient who is moving for a longer period of time is not likely to be in danger due to a serious apnea and, therefore, the motion information itself is giving information about the patient, e. g. the motion could be also linked to the discomfort of the infants [170].

Our motion classification reached an accuracy equal to 88.22% in the training and testing set. It should be noted that the accuracy results are underestimating the real accuracy. The manual annotation was performed by a single person and while the automatic classification is on a second by second basis (due to the sliding window's slide), the manual annotation tends to overlook particularly short events. An example is visible in Figure 5.8, the detected label (the result of the automatic classification) can present fast oscillations, whereas the manual annotation is more stable and sometimes stretched compared to the detected label (anticipated starting point and/or postponed ending point). The validation set obtained a lower accuracy result, i. e. 82.52%, this is due to the reduced sensitivity of our motion classification on this dataset. These results could indicate that not enough data were included in our optimization step or that the training dataset is not representative enough. Differences were observed between the two sets in the blanket position, which could end up hiding some of the moving pixels. Whether the infant's sleeping position plays a role warrants further analysis. Moreover, the motion detection strategy, as it is implemented now, is limited by changes in the distance (between camera and infant) or zoom, however, all infants in our study occupy a similar portion of the image, although small variations are present. The method may need to be optimized for different distances or features in the images could be used to make the method independent of the distance.

Table 5.4 presents a comparison of the MAE obtained by our previously published method [169], detailed in Chapter 4, and the adjusted one presented in this chapter, obtaining an improvement of around 1 BPM on the average MAE. The harmonic problem was particularly noticeable in one of the infants, i. e. infant 8, where the introduction of our adjustment drastically reduced the error (from 7.17 to 1.89 BPM). The NNS is present in less than 5% of the included segments. This is mostly due to the study protocol since hours in which the parents were not in the wards were preferred, as the babies would then spend more time in the bed, but this was not always possible. The percentage of presence of NNS is, therefore, likely underestimated and not completely representative. This percentage could be higher if the parents are in the neonatal ward next to the infant or in home-care because the soother will be given to the baby more often in these cases. The MAE obtained during NNS is reduced in our new implementation, though still higher than the average MAE considering all usable segments. NNS frequencies have been reported to vary and can correspond to the ones of the normal RR or be higher up to 150 sucks per minute [171, 172]. Therefore, if the frequencies of NNS are higher than the normal RR range, our algorithm can detect the respiration pixels and correctly estimate the RR. However, if the NNS frequency is inside the respiration band, our method can no longer discriminate between NNS and respiratory signals. This is a limitation present in all methods that automatically identify the region of interest (ROI) or technologies that monitor the motion in an area, e. g. continuous-wave radars. This problem, particularly important for home-care and babies cared for in open beds, should be further analyzed.

Furthermore, in Table 5.5, a comparison of the results between the training and testing set, and the validation set, in the usable moments excluding NNS, is provided. The errors are higher in the validation set compared to the training and testing one. We believe this is a consequence of the reduced sensitivity of the motion classification step for the validation set, which leads to the inclusion of segments with type 1 motion in the moments used for the estimation of the RR. Other factors influence the average error, one is the presence of babies breathing with a periodic breathing (PB) pattern, a physiological breathing pattern in infants associated with the alternation of normal breathing and breathing pauses [18]. One of the babies in the training and testing set continuously breathes following a PB pattern (infant 3), whereas another baby in the training and testing set (infant 4) and six babies in the validation set (infants 10, 11, 12, 13, 15, and 17) resulted in having segments with a PB pattern. PB pauses have been reported to last 6 to 9 s [132], in our dataset, we observe breathing pauses with a duration of up to 10 s. It becomes evident that by using a window size of 8 s, we will detect an RR in windows that do not contain any respiration-related oscillations. This causes the estimation of the error to be higher than the real one because both our method and the CI will provide an incorrect estimation of the RR, an example is visible in Figure 5.9. Our method requires the selection of respiration pixels in every window, if there is no respiration information in the video segment, the selected pixels will contain noise. The results are, therefore, also dependent on the length of the breathing pauses which can be different for each baby. This problem needs to be further analyzed considering also apneas, and the number of pixels selected could be used as an indicator to detect the absence of respiration. The PB pattern is, anyway, clearly visible in the time signals, and in the future, methods for cessations of breathing detection such as [43, 173] could be used to identify the breathing pauses and remove these from the RR comparison. Moreover, some of the recordings in the validation set (belonging to infants 13 and 14) contain segments in which the respiration motion is not visible due to the blanket position,

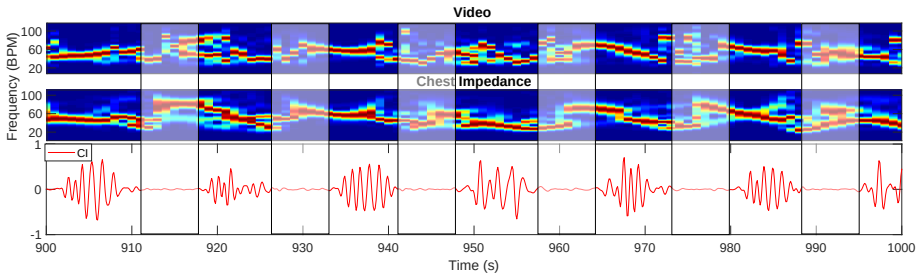


Figure 5.9: Example of results with PB. The sudden changes in RR can be seen in the STFTs close to the breathing pauses (indicated using the rectangular boxes with width of 8 s).

directly influencing the error. This problem was highlighted also in Chapter 4.

By comparing the errors in the ideal moments where the infants are still and in the moments where type 2 motion occurs, differences can be noted. On average, the MAE during type 2 motion segments is higher than the one during ideal moments, with an increase of 1.16 and 1.97 BPM for training and testing, and validation set, respectively. We believe the cases in which the errors are higher for the type 2 motion may be related to the position of the pixels containing respiration. Our approach is based on the assumption that respiratory pixels are visible on the edge of the blanket and chest/neck area, and type 2 motion, like arm motion or head motion, will not affect our performance. However, this is not always true, like in cases where most of the respiration pixels come from the arm or the head itself, which is happening in some babies' videos. This is again caused by the blanket covering the main source of respiration signal, i. e. the chest. We can expect this problem to be further reduced in infants in incubators that are not covered. The inclusion of the type 2 motion segments allows to drastically increase the amount of time used for respiration estimation at a cost of a higher error.

The two videos of infants 8 and 9 collected using an RGB camera seem to perform better, yielding lower MAE compared to the other babies (except for infant 1 whose MAE is comparable). However, we believe that conclusions regarding which technology performs best cannot be drawn from this comparison, as such would require a dataset acquired simultaneously with both camera types. Moreover, the RGB videos were not included in the validation set, therefore, the performance of our algorithm on this type of video should be further analyzed and more data should be included.

Overall, our MAEs and Bland–Altman plots are comparable with studies performed in similar populations, e. g. the work of Villarroel et al. [79] showed a MAE of 4.5 and 3.5 BPM for their training and test set respectively, very similar compared to our 3.31 and 5.36 BPM. Our method, though, can be used on both thermal and RGB/NIR cameras, provides motion information, and does not rely on skin visibility but only on respiration motion being visible. The limits of agreement in Figure 5.6b are higher than the ones in the training and testing set and higher compared to the results obtained in [79], this is due to a combination of the problems previously described.

This chapter provides promising results and highlights possible challenges for neonatal respiration monitoring. In particular, in the cases of babies cared for in an open bed and babies in a home-care environment, the NNS presence and its effect on unobtrusive vital signs solutions should be investigated further, although the presence of the NNS motion itself could indicate the absence of critical situations. Moreover, one of the main limitations of our method, but in general of camera-based solutions, is the respiration motion being hidden by blankets covering the infants. While camera-based solutions provide contextual information undoubtedly usable for the detection of motion, they may also require the fusion with a different technology that would not be affected by this type of problem, such as

radar or pressure-sensitive films, or a clearer protocol for blanket positioning.

5.5 Conclusions

This chapter presents a combination of a method for motion detection, optimized to detect motion hiding respiration, and a method for RR detection that, using three features, automatically selects the pixels of interest. The motion robustness achieved thanks to our features, allows us to increment the amount of time used for camera-based respiration detection, including segments that contain limbs or head movements. The test of the cross-validation obtained an accuracy of around 88% in the motion identification. A lower accuracy was obtained in our validation set, indicating that the optimization could be improved. The RR estimation was compared with the CI reference and yielded an average MAE of 3.31 and 5.36 BPM for the training and testing set, and validation set, respectively. The MAE during type 2 motion was higher than the one in the ideal moments of 1.16 and 1.97 BPM for the training and testing set, and validation set, respectively. This proves the motion robustness is improved, but more work is needed to achieve continuous unobtrusive respiration monitoring. Limitations on the use of camera-based solutions in a neonatal ward environment are highlighted in this chapter, i. e. the PB influence on the errors, the blanket covering respiration motion, and the NNS presence. This method can be used for different camera modalities and does not require skin visibility.

CHAPTER 6

Automatic separation of respiratory flow from motion in thermal videos

Abstract

Both respiratory flow (RF) and respiratory motion (RM) can be visible in thermal recordings of infants. Monitoring these two signals usually requires landmarks detection for the selection of a region of interest. Other approaches combine respiratory signals coming from both RF and RM, obtaining a mixed-respiratory (MR) signal. The detection and classification of apneas, particularly common in preterm infants with low birth weight, would benefit from monitoring both RF and RM, or MR, signals. Therefore, we propose in this chapter an automatic RF pixels detector not based on facial/body landmarks. The method is based on the property of RF pixels in thermal videos, which are located in areas with a smooth circular gradient. We defined 5 features combined with the use of a bank of Gabor filters that together allow selecting the RF pixels. The algorithm was tested on thermal recordings of 9 infants amounting to a total of 132 minutes acquired in a neonatal ward. On average the percentage of correctly identified RF pixels was 84%. Obstructive apneas (OAs) were simulated as a proof of concept to prove the advantage in monitoring the RF signal compared to the MR signal. The sensitivity in the simulated OA detection improved for the RF signal reaching 73% against the 23% of the MR signal. Overall, the method yielded promising results, although the positioning and number of cameras used could be further optimized for optimal RF visibility.

This chapter is based on:
I. Lorato, S. Stuijk, M. Meftah, D. Kommers, P. Andriessen, C. van Pul, and G. de Haan, "Automatic separation of respiratory flow from motion in thermal videos for infant apnea detection," Sensors, vol. 21, no. 18, pp. 1–16:6306, 2021.

6.1 Introduction

Respiration is one of the most important vital signs, able to detect early clinical decline [174]. It can be monitored in hospital wards to detect critical events and respiratory irregularities. In neonatal intensive care units (NICUs), in particular, the immaturity of the respiratory control system of premature infants is the main cause of apnea of prematurity (AOP), which is one of the most common diagnoses [115]. Infants' breathing patterns can present cessations of breathing (COBs), the ones that last 20 seconds or 10 seconds accompanied by bradycardia and/or desaturation are called apneas [116]. Three main categories of apneas can be defined: central apnea (CA), which is characterized by cessation of both respiratory flow (RF) and effort, obstructive apnea (OA) which is caused by the collapse of the upper airway and it manifests as cessation of airflow and presence of respiratory effort, and mixed apnea (MA) which is a mixture of the previous two [115].

Infants in NICUs are typically monitored using several adhesive skin sensors and electrodes. Respiration is monitored using chest impedance (CI), which is measured using the electrocardiogram (ECG) electrodes. However, this method is not able to detect OAs, due to the presence of respiratory effort [175]. Discriminating between different types of apnea is difficult due to limitations of the technology used. Moreover, both CI and other methods that may be used to improve the detection of apneas require attaching sensors and electrodes to the skin or positioning sensors close to the nostrils. This causes discomfort or even skin irritation in preterm infants. Because of these reasons, unobtrusive monitoring of vital signs, and in particular of respiration is being investigated.

Numerous technologies are being researched for the unobtrusive monitoring of respiration: radars [124, 176], red-green-blue (RGB) and near-infrared (NIR) cameras [79, 177], thermal cameras or infrared thermography (IRT) [103, 110], pressure-sensitive films [66, 158], and vision systems with depth sensing [178, 179]. All these techniques, apart from thermal cameras, are able to monitor only respiratory motion (RM). The clear advantage in monitoring both RF and motion is the more accurate detection and classification of apneas. Identifying the type of apnea is clinically relevant as the required therapeutic intervention can be different [13]. Studies suggest that the phase shift between abdomen motion and thorax motion can also be monitored using normal RGB/NIR cameras and could be used to identify OAs [83]. However, the identification of the regions can be quite challenging, especially for infants or in general if the subject is covered by a blanket or sheet. IRT remains the most promising option when aiming at unobtrusive apnea detection and classification.

Studies using thermal cameras for the detection of RF have been published in recent years. Several works proposed to detect RF based on a manually selected region of interest (ROI) [99–101,109–112], others used facial and/or body landmarks based solely on thermal images [91, 95, 97, 99] or by combining it with RGB/NIR

images [104, 106, 114]. While facial and body landmarks detection can be used in controlled settings, using them in hospital settings without any constraints on the subjects' position is quite challenging [102]. Moreover, the identification of multiple ROIs is required, considering that patients could be nose or mouth-breathing. In addition, by selecting ROIs specifically at the nose/mouth areas the thermal variations due to RF which can be registered in the environment, e. g. on the pillow, would be ignored. This is, sometimes, the only source of RF in the video, as will be shown in this chapter. Recent solutions propose to automatically identify respiratory pixels in thermal videos [102, 169, 180], also described in Chapters 4–5. These methods, however, have the disadvantage of mixing the thermal variations due to RF and motion into a single signal, obtaining a mixed-respiratory (MR) signal, the terminology used is further detailed in Section 6.2.

For these reasons, we propose in this chapter an automatic RF pixels detector using characteristics that set RF and RM pixels apart. Particularly, RF pixels can be located in non-edge areas of the videos and due to the thermal diffusion the airflow generates areas with a smooth gradient. Also, RM can be in-phase or in anti-phase with the RF. The use of this algorithm results in an RF signal, which can be used for a more accurate detection and classification of apneas. The algorithm will be applied to a set of low-resolution infants' thermal videos collected in a neonatal ward. The data amounts to a total of 132 minutes split between 9 infants. Finally, we adapted the videos to simulate OAs and used these as a proof of concept for apnea detection using the COB-detector described in Chapter 2. The performance in the apnea detection was analyzed for the different respiration signals obtainable from a thermal video. Particularly, we compared the performance of the RF signal, obtained with the method proposed in this chapter, with the one of an MR signal which is obtained using the method explained in Chapter 5 and in [180].

Section 6.2 details the background and reasoning behind the method, and describes the processing steps and the experimental setup. Section 6.3 and Section 6.4 contain the results and the discussion, and Section 6.5 provides the conclusions.

6.2 Materials and methods

In thermal videos, respiration can be detected by monitoring RF, RM, and/or both. The videos, therefore, contain RM pixels, RF pixels, and noise-pixels. Pixels simultaneously affected by RF and RM can also be present, the term MR pixels is used to indicate the pixels belonging to this category as well as the ones belonging to the RF or RM pixels. Therefore, we define:

- *RF signal*: the signal obtained by combining together only the RF pixels.
- *RM signal*: the one obtained using only the RM pixels.

- *MR signal*: this signal can be obtained by combining respiration pixels, regardless their origin, i. e. the MR pixels. This is the signal that was obtained in the previous Chapters 4–5, and earlier research [102, 169, 180].

Aiming at apnea detection and classification, we need to monitor the *RF signal* since this allows us to accurately detect OAs. However, to differentiate between an OA and a CA, motion information is also needed. It could be possible to monitor the *RM signal*, but the *MR signal* can also be used as it will become a *RM signal* in the windows containing apneas (since the flow contribution will not be present in those segments). Based on this reasoning we aim at monitoring the *RF signal* and the *MR signal* which can potentially ensure an accurate apnea classification and detection.

To obtain the *RF signal*, a preferably automatic selection of the RF pixels is needed. To arrive at an automatic detection, we may use the following characteristics:

1. RM pixels are located at (typically sharp) edges in the thermal image, e. g. the boundary of the head. Without a gradient, the RM would not be visible. Moreover, the steeper the gradient, the stronger the temporal signal. RM pixels typically extend 1-dimensionally (along an edge).

2. RF pixels can be located in non-edge areas of the image. Moreover, the temperature changes due to ex/inhalation generate areas with a smooth circular gradient, caused by thermal diffusion. Consequently, they typically extend 2-dimensionally in the image.

3. *RM signals* can be in phase, or in anti-phase with the *RF signal*, depending on the direction of the motion and the temperature gradient. The RF always results in warming regions during exhalation and cooling regions during inhalation, whereas the RM, for example, could be visible at the edge between blanket and face resulting in warm pixels becoming colder during inhalation, or at the edge between the infant's head and the sheet, resulting in colder pixels becoming warmer during inhalation.

6.2.1 Materials

The videos used in this chapter were collected on infants who were nursed in an open bed in the Neonatal Medium Care Unit of the Máxima Medical Centre (MMC) in Veldhoven, The Netherlands. The study received a waiver from the ethical committee of the MMC and the infants' parents signed an informed consent before the study.

Dataset

The thermal videos collected amount to a total of around 42 hours acquired from 15 infants. A manual annotation was performed to analyze the content of the videos. All events visible in the videos were annotated. Particularly, the movements of the infants were divided into three main categories, still, type 1 motion, and type 2 motion. Still is when all body movements are absent apart from RM, type 1 and 2 motion were defined to differentiate between movements including the chest and movements involving only other parts of the body. Also, the presence of a soother was annotated, as this may affect the algorithm. Caregivers' and parents' interventions, baby out of bed, motion of the camera, and unsuitable camera view were also annotated. More details on the annotation can be found in Chapter 5.

In this chapter we focus solely on the video segments in which the infants are still, all other events were neglected. The total absence of body motion was preferred to allow an accurate annotation of the RF pixels location, detailed in the next section. Moments in which the infant was still but had the soother were also not considered since the soother may end up physically hiding RF. The moments in which the infants were still for at least 30 s amount to 339 minutes considering all the 15 infants. Segments shorter than 30 s were not used.

To eliminate video fragments without visible flow, an annotation was carried out by watching the unprocessed videos. The main strategy was looking for pseudo-periodic temperature variations at nostrils or textiles around the infant. Knowing that RF was present in the video segments was required for the development and testing of the algorithm. This selection results in around 142 minutes unequally split on 11 infants. Moreover, two babies presented periodic breathing (PB), a benign breathing pattern made of an alternation of COBs and normal breathing [14]. These infants were removed from the usable data since the presence of COBs, not annotated, would cause underestimating the performance of the RF pixels selection and of the OA detection. Therefore, our dataset in this chapter amounts to around 132 minutes split in 87 segments and unevenly distributed on 9 infants. Figure 6.1 presents the distribution of the data and the percentage on the total 42 hours. Table 6.1 shows the infants' information, all the infants are in supine position except Infant 4 who is in prone position and Infant 1 is the only infant with a nasal cannula. Moreover, Infant 2 and Infant 6, are the only infants in this dataset that do not have a nasogastric feeding tube. For further comments about the flow visibility in the recordings, refer to Section 6.4.

Annotation of the respiratory flow pixels location

To obtain a reference *RF signal* and to be able to estimate if our algorithm selects the correct pixels, each video segment of each infant was examined to locate the RF pixels. MATLAB (MATLAB 2018b, The MathWorks Inc., Natick, MA, USA) was used for the manual annotation, and bounding boxes were used to annotate all the frame regions affected by airflow. Regions containing flow may occur at the

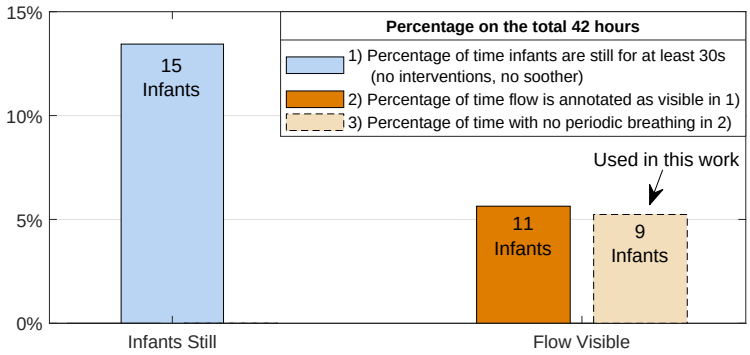


Figure 6.1: Percentages of video segments in which the infants are still with a minimum duration of 30 s, with flow visible, and with no PB. All percentages are reported with respect to the total recording time of 42 hours.

nostrils, on the mouth, and/or on textiles surrounding the infants' face, these were all considered as valid RF pixels positions. The RF pixels were annotated at the location where pseudo-periodic temperature variations were visible in the thermal videos. Figure 6.2 shows the relative importance of the areas annotated as RF pixels in our videos. The figure also shows the individual percentage per infant of segments with flow visible compared to segments with the infant still for at least 30 s. The annotated RF pixels were used as the ground truth to evaluate the performance of the automatic RF pixels detection. A *reference respiratory flow (RefRF) signal* was obtained by averaging all the annotated RF pixels and was

Table 6.1: Information on the infants and the data used in this chapter. The gestational age (GA) is indicated in weeks (w) and days (d), postnatal age (PA) is in days.

Infant	GA (w+d)	PA (days)	Total Duration (minutes)	Number of Video Segments
1	26w 4d	59	4.84	4
2	38w 5d	3	1.86	3
4	26w 3d	59	26.93	28
6	40w 1d	6	17.53	11
10	26w 4d	77	24.48	16
11	26w 4d	77	8.29	4
14	32w 2d	11	3.62	3
15	35w 1d	8	33.84	16
17	27w 5d	16	10.78	2

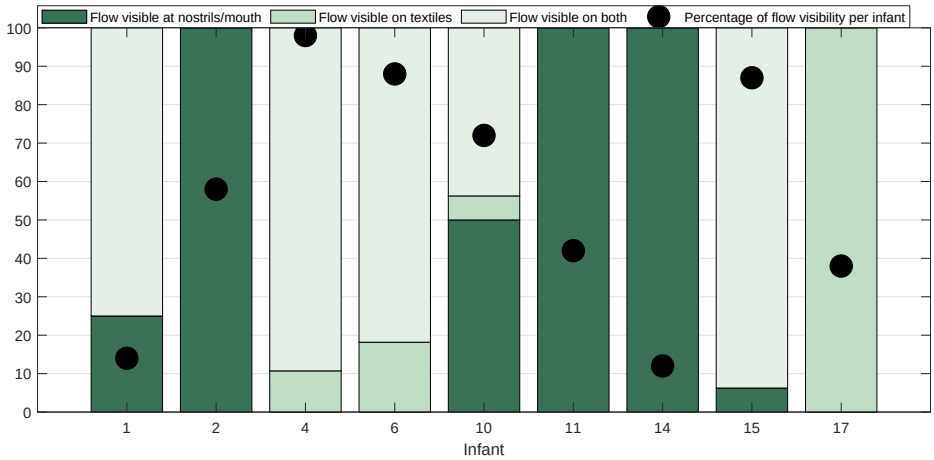


Figure 6.2: Information on the location of the annotated RF pixels for each infant and the percentage per infant of segments with flow visible compared to the segments annotated as still for at least 30 s.

used for comparison purposes with the *RF signal* obtained using our algorithm.

Obstructive apnea simulation

Our dataset did not include naturally occurring OAs, however, we wanted to include a proof of concept of enhanced OA detection with RF monitoring. Therefore, we built a dataset with simulated OAs, using the earlier videos. In total, we simulated 87 OAs. To simulate an OA, we substituted the annotated RF pixels with noise-pixels from the images for a time period of 10 s in the middle of each video segment. The low-frequency content of the noise-pixels was removed and replaced with the low-frequency content of the original RF pixels. In this way, all RF pixels will contain noise for the selected 10 s, and the RM pixels are, instead, not altered.

Experimental setup

The videos were collected using three FLIR Lepton 2.5 cameras, these devices were chosen based on their relatively low cost, which allows the use of multiple cameras improving the coverage of the patient. The cameras are sensitive in the long-wave infrared (LWIR), have a resolution of 60×80 pixels, and an average frame rate equal to 8.7 Hz. The three cameras were positioned around the infants' open bed. The CI from the patient monitor (Philips MX800) was acquired as the reference respiratory signal. For more information regarding the setup, we refer to Chapter 4.

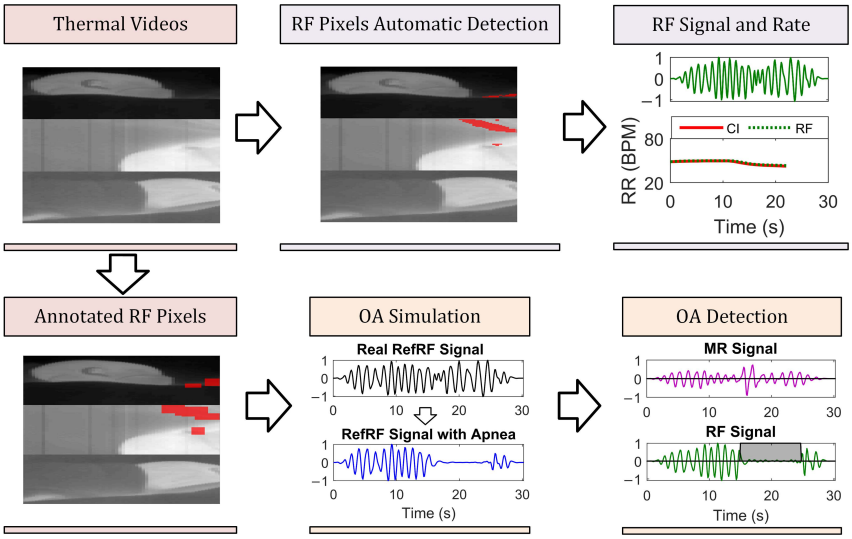


Figure 6.3: Summary of the processing and example results. RF pixels are automatically detected in the thermal videos and used to calculate the *RF signal* and the flow-based RR. Moreover, the RF pixels location is manually annotated. The annotated RF pixels are substituted with noise to simulate the occurrence of an OA. A COB-detector is used to compare the performance in OA detectability between the *RF signal* and the *MR signal*.

6.2.2 Methods

The identification of the RF pixels is based on five features, partially already introduced in [180] and in Chapter 5, combined with the new use of a bank of Gabor filters. These filters allow us to exploit the characteristic of RF pixels, i. e. that they typically occur in 2D-smooth areas. Using the chosen pixels, a *RF signal* and the flow-based respiration rate (RR) were obtained. Moreover, OAs were simulated in the 87 video segments and our previously proposed method for the detection of COBs [173] was used. The method has been detailed in Chapter 2. This allows comparing the detectability of the OAs in the different respiration signals we obtain from the thermal videos (i. e. *MR signal* and *RF signal*). These steps are summarized in Figure 6.3.

Preprocessing

The thermal images coming from the three camera views were merged on the same image plane as explained in Chapter 4 and as visible in Figure 6.3, obtaining a single video with resolution 180×80, i. e. $M \times L$. Each pixel time-domain signal was interpolated with a 1D linear interpolation to compensate for the uneven sampling

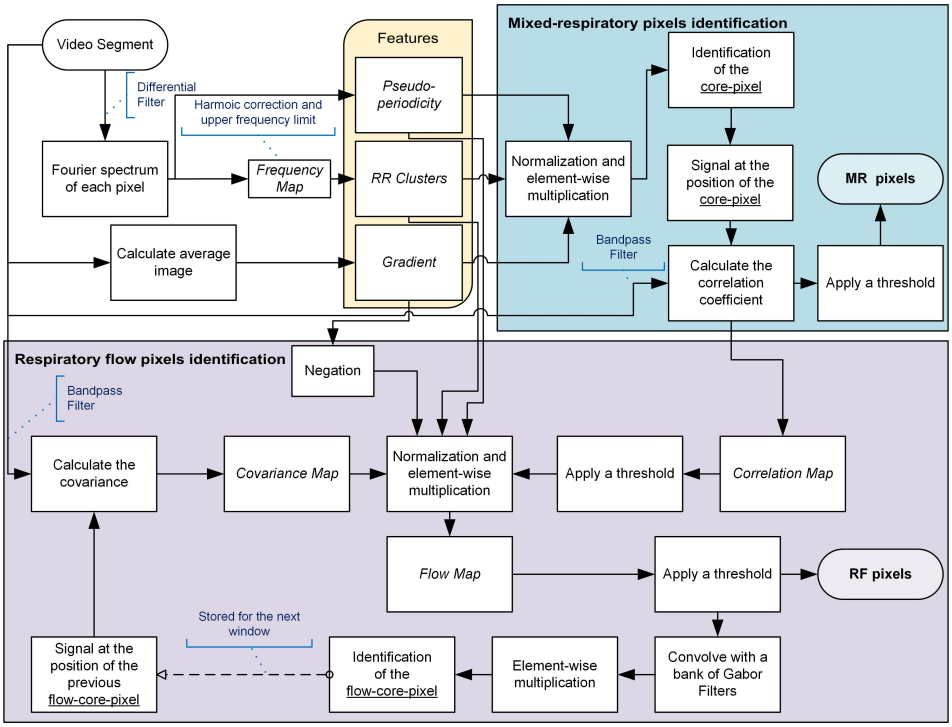


Figure 6.4: Simplified block diagram of the respiratory pixels selection algorithm, for both MR and RF pixels.

rate. The resulting frame rate is 9 Hz, close to the average frame rate of the FLIR camera. This was also explained and used in our previous works [169, 180] and in the previous Chapters 4–5.

Respiratory flow detection

The method for the automatic detection of RF pixels, is based on the aspects previously explained. Briefly, a set of 5 features is combined to identify the RF pixels. A flow-core-pixel, i. e. a pixel that is most likely to belong to the RF pixels, is selected as it will be used as a basis for the calculation of one of the features. Gabor filters are introduced for the accurate selection of the flow-core-pixel. A simplified block diagram of the algorithm for the detection of both MR pixels and RF pixels is shown in Figure 6.4. The time-domain signals of each pixel in each window are referred to as $x_{m,l}(nT_s)$, where (m, l) indicates the pixel position, and $n = 0 + (j-1)/T_s$, $1 + (j-1)/T_s$, ..., $(N-1) + (j-1)/T_s$. Each window is identified by the integer j, and consists of $N = 72$ consecutive samples in an 8 s fragment, sliding in steps of 1 s. The sampling period T_s equals 0.111 s.

Gabor filters are well-known bandpass filters used in image processing for texture and edge detection. The kernel is formed by a sinusoidal carrier and a 2D Gaussian envelope. Several Gabor filters can be generated by varying the spatial frequency of the sine wave and the orientation of the filter. By applying a set of filters to an image, edges and textures can be emphasized. Considering the properties of the distribution of RF pixels and RM pixels, we apply a bank of Gabor filters by varying the orientation and the spatial frequency aiming at locating RF pixels, which should have a similar response for all orientations. For RM pixels, on the other hand, we expect a higher response in specific directions, being mostly along, possibly curved, lines. We used the MATLAB built-in function to generate the bank of Gabor filters, i. e. *gabor*. We empirically selected a set of parameters for the orientation and for the spatial frequency of the filters, $\lambda = 3, 4, ..., 8$ pixels/cycle and $\theta = 10°, 20°, 30°, ..., 170°$. Multiple spatial frequencies were chosen to allow the method to work with both flow visible at nostrils/mouth or flow visible on textiles, as these usually produce regions affected by flow with different sizes. The bank is applied to an input map called *Flow Map*, which will be defined later, by convolving the input map with each Gabor filter. In particular:

$$\Psi(\lambda, \theta) = |\widehat{\mathbf{FM}} \otimes \Gamma(\lambda, \theta)|, \tag{6.1}$$

where $\Gamma(\lambda, \theta)$ represent a Gabor filter, and $\widehat{\mathbf{FM}}$ is the input map. The $\Psi(\lambda, \theta)$ are the magnitudes of the Gabor responses for each spatial frequency λ and orientation θ. We select the flow-core-pixel by multiplying all the Gabor responses $\Psi(\lambda, \theta)$. The flow-core-pixel is the pixel corresponding to the highest value in the image resulting from the multiplication:

$$(m_{p_f}, l_{p_f}) = \underset{(m,l)}{\arg\max} \left(\prod_{\lambda,\theta} \Psi(\lambda, \theta) \right). \tag{6.2}$$

Therefore, (m_{p_f}, l_{p_f}) indicates the position of the flow-core-pixel in each window. The map given as input to the Gabor filters is called *Flow Map* and is a combination of 5 features. In particular:

$$\mathbf{FM} = \hat{\mathbf{C}} \cdot \tilde{\mathbf{C}}_{flow} \cdot \tilde{\mathbf{Q}} \cdot \tilde{\mathbf{W}} \cdot (\mathbf{J} - \mathbf{G}). \tag{6.3}$$

$\tilde{\mathbf{C}}_{flow}$ is a new feature introduced to locate RF pixels more accurately, and called *Covariance Map*. Each element represents the covariance between the signal of the flow-core-pixel found in the previous window and the signal of the other pixels in the video segment. $\tilde{\mathbf{C}}_{flow}$ is the normalized version of the *Covariance Map*:

$$c_{flow_{m,l}} = \begin{cases} 1 & \text{if } j = 1 \\ \frac{1}{N} \sum_{t=1}^{N} \hat{x}_{(m_{p_f}, l_{p_f})_{j-1}}(t) \cdot \hat{x}_{m,l}(t) & \text{otherwise.} \end{cases} \tag{6.4}$$

$c_{flow_{m,l}}$ represents, therefore, the covariance between the signal of the chosen flow-core-pixel in the previous window $(m_{p_f}, l_{p_f})_{j-1}$ and the signal of a pixel in position

(m, l), $\hat{x}_{m,l}(t)$ and $\hat{x}_{(m_{p_f}, l_{p_f})_{j-1}}(t)$ are the filtered time-domain signals, whereas t is an index that sweeps through the samples in the jth window. The time-domain signals were filtered with a bandpass filter between 30 and 110 breaths per minute (BPM), i. e. the expected breathing frequency range of an infant. The $c_{flow_{m,l}}$ were then normalized resulting in a matrix between -1 and 1, i. e. $\tilde{\mathbf{C}}_{Flow}$. The covariance was preferred to the correlation coefficient because it allows taking into consideration also the standard deviation of the time signals, which is advantageous assuming the biggest thermal variations are associated with respiration. Moreover, the sign of the covariance was kept which allows rejecting anti-phase signals, which can only originate from motion.

The other 4 features in Equation 6.3 were previously developed to obtain a *MR signal* from thermal videos, for a detailed explanation refer to [180] and Chapter 5. These features were designed to locate MR pixels but can be adapted for the identification of the RF pixels. $\mathbf{Q}$ is called *Pseudo-periodicity* and is based on the estimation of the height of the normalized spectrum's peak. $\mathbf{W}$ is called *RR clusters* and is based on the application of a 2D non-linear filter for the detection of pixels that have similar frequencies nearby. $\mathbf{G}$ is *Gradient*, which identifies the edges of the thermal images. These three features were used to identify a core-pixel, i. e. a pixel that best represents the *MR signal*. Once a core-pixel is found the Pearson correlation coefficient is used to locate all the other pixels containing respiration signals and the *MR signal* is obtained by averaging these pixels together as explained in our previous work [180] and in Chapter 5. The Pearson correlation coefficients obtained between the core-pixel and all the other pixels are arranged in a matrix called *Correlation Map* and indicated with $\mathbf{C}$. The *Correlation Map* obtained from the core-pixel can be used to locate the MR pixels. We binarized this map by applying an empirical threshold κ_3 on the absolute values equal to 0.6:

$$\hat{\mathbf{C}} = |\mathbf{C}| > \kappa_3. \tag{6.5}$$

The *Flow Map* can be obtained by combining this binarized *Correlation Map* with the *Covariance Map* and the other features as explained in Equation 6.3. $\tilde{\mathbf{Q}}$, $\tilde{\mathbf{W}}$, and $\mathbf{G}$ represent the *Pseudo-periodicity*, *RR clusters*, and *Gradient* features respectively, the tilde is used to indicate that the features were normalized between 0 and 1, $\mathbf{G}$ is already binary. $\mathbf{J}$ is an $M \times L$ matrix containing all ones and, therefore, the combination with the *Gradient* feature gives a weight equal to 1 to the non-edge regions. The *Flow Map* was then binarized by applying an empirical threshold, ξ_1 equal to 0.2:

$$\widehat{\mathbf{FM}} = \mathbf{FM} > \xi_1. \tag{6.6}$$

Even though the combination of these features allowed removing most of the RM pixels from the selectable pixels, in the first window the *Covariance Map* is not computed and some of these pixels may still be present in the binarized *Flow Map*. Moreover, considering the flow-core-pixel is used to estimate the *Covariance Map* in the following windows, the detection of the right pixel is particularly important.

Additionally, the *Flow Map* may still contain some noise-pixels as well as the flow ones. Therefore, to select the flow-core-pixel accurately, we introduced the bank of Gabor filters, and the $\widehat{\mathbf{FM}}$ was given as input to the bank, as done in Equation 6.1.

The RF pixels are, therefore, detected, in the first window of each video segment, only the flow-core-pixel is used, afterwards, all non-zero pixels in $\widehat{\mathbf{FM}}$ are considered RF pixels:

$$\mathbf{P}_{flow} = \begin{cases} (m_{p_f}, l_{p_f}) & \text{if } j = 1 \\ \\ (m, l) \ : \widehat{\mathbf{FM}}(m, l) = 1 & \text{otherwise.} \end{cases} \tag{6.7}$$

$\mathbf{P}_{flow}$ is, therefore, a set containing the positions of the detected RF pixels. The conditions for RF pixels detection are quite strict, it could happen that no pixel is found, in that case the previously chosen RF pixels are used in the current window as well. The *RF signal* is obtained by averaging together all RF pixels contained in $\mathbf{P}_{flow}$. An example of the features in the first window is shown in Figure 6.5a, and the features obtained in the following window in Figure 6.5b. The figures show the advantage introduced by the use of the *Covariance Map*, rejecting anti-phase RM pixels. As a consequence, the *Flow Map* in Figure 6.5b does not contain RM pixels compared to the *Flow Map* in Figure 6.5a.

The *MR signal* is also obtained from the videos, using our method previously described in [180], and will be used for comparison purposes.

Obstructive apnea detection

We adjusted our previously published COB-detector [173], presented in Chapter 2, to evaluate the detectability of OAs, which were simulated as previously described. The COB-detector assumes that COBs can be detected by monitoring sudden amplitude changes and it is based on the comparison of a short-term standard deviation and a long-term standard deviation. The only adaptations applied to our previous published implementation concern the length of the windows for the calculation of the two standard deviations. The duration of these windows was chosen in Chapter 2 based on the targeted COB. In particular, the window for the calculation of the short-term standard deviation should be close to the minimum COB duration, an apnea of 10 s. The window for the long-term standard deviation, which is calculated as the median of the short-term standard deviations, must be higher than the COB duration. Otherwise, the long-term standard deviation will dynamically adapt to the standard deviation during the apnea event (i. e. detecting the cessation of the event while the apnea is still ongoing). In our current implementation, the short-term standard deviation is calculated using 8 s windows, which is the same sliding window approach used for the RR estimation. The long-term standard deviation is calculated in a window of 15 s. This window could be reduced to 11 s considering the fact that this is closer to the designed duration of the OA, but we kept it higher to easily adapt to non-simulated cases.

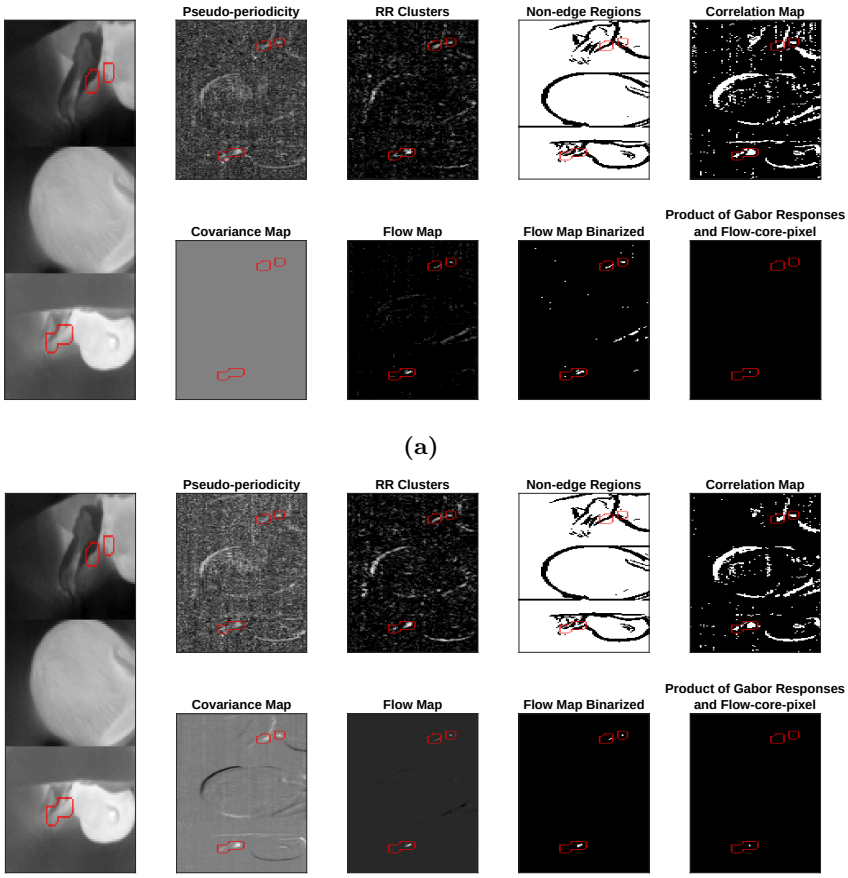


Figure 6.5: Example of features used for the detection of the RF pixels, the location of the annotated RF pixels is indicated with a red perimeter: **(a)** the features and the choice of the first flow-core-pixel; **(b)** the features used to locate the RF pixels in the next window.

The *RF signal* and *MR signal* were obtained as described in the previous section using our dataset with simulated OAs. The COB-detector was applied to the *RefRF signal*, as reference of the results achievable when monitoring RF, on the *RF signal* obtained from applying our method, and also on the *MR signal* to highlight the limitations of monitoring this type of signal when aiming at apnea detection.

Evaluation metrics

To evaluate the performance of the RF pixels detector, we used the annotated RF pixels as a reference. In each window, we evaluate the percentage of detected RF pixels that belong to the annotated RF pixel areas. In particular:

$$PF(j) = \frac{\#(\mathbf{P}_{flow} \bigcap \mathbf{P}_{ann})}{\#(\mathbf{P}_{flow})} \cdot 100, \tag{6.8}$$

the symbol # is used to indicate the cardinality of the sets, and $\mathbf{P}_{ann}$ is a set containing the annotated RF pixels. Moreover, to estimate the number of RM pixels erroneously included in the $\mathbf{P}_{flow}$ set, we calculate the percentage of the detected RF pixels that belong to the pixels used to calculate the *MR signal* after removing $\mathbf{P}_{ann}$ from the set. Formally:

$$PM(j) = \frac{\#(\mathbf{P}_{flow} \bigcap (\mathbf{P}_m - \mathbf{P}_{ann}))}{\#(\mathbf{P}_{flow})} \cdot 100, \tag{6.9}$$

with $\mathbf{P}_m$ indicating the set of pixels used to obtain the *MR signal*. The $PF(j)$ and $PM(j)$ are then averaged to obtain an average percentage of correct and incorrect pixels detected in each video segment. The mean absolute error (MAE) is also estimated to compare RRs obtained using the *RF signal*, the *MR signal*, or the *RefRF signal* and the RR of the CI reference.

For the OA detection step, accuracy (ACC), sensitivity (SE), and specificity (SP) are calculated, by comparing the OA detection result with a template signal. The template signal has been built to be equal to one in the segment containing a simulated OA, and to zero in the rest of the signal. ACC, SE, and SP are calculated as defined in [71,173], i. e. by considering the total duration of the time intervals with OAs correctly and incorrectly detected (time true positive and time false positive), and correctly and incorrectly not detected (time true negative and time false negative).

6.3 Results

Figure 6.6 shows an example for each infant of the detected and the annotated RF pixels. This figure clearly shows the variability in flow location and infant positions contained in our relatively small dataset. Note that each video segment of each infant may have a different flow location, due to the infant moving, therefore, this example figure does not cover all the cases present in the dataset.

The results of the RF pixels detection step and the MAE obtained for the different respiration signals obtained are shown in Table 6.2. The percentage of correct RF pixels detection, PF, is on average equal to 84.28%, and PM, RM pixels erroneously included, is on average 0.35%. The average MAEs obtained by comparing the RRs of the CI reference with the one of *RF signal*, *RefRF signal*,

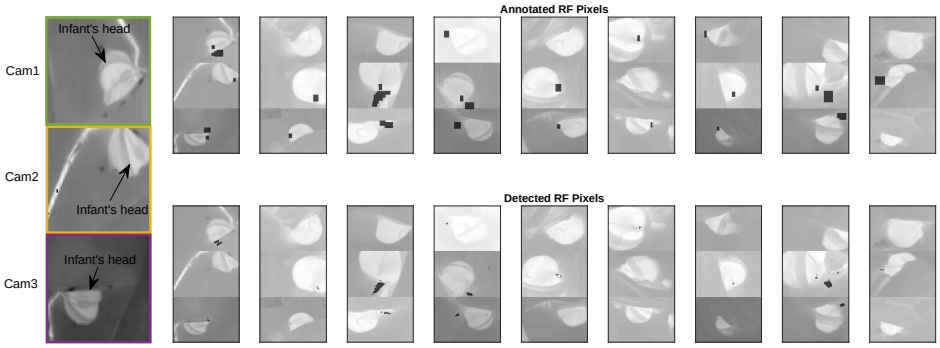


Figure 6.6: An example thermal image with a description of the content, and examples of the annotated RF pixels and the detected ones for the different infants included.

and *MR signal* are respectively 2.20 BPM, 1.85 BPM, and 2.11 BPM. Moreover, Table 6.3 contains the results of the OA simulation and detection step, showing ACC, SE, and SP calculated using the three different signals obtained from the videos. The superior performance of the *RF signal*, average SE 73%, compared to the *MR signal*, average SE 23%, in detecting the simulated OA is clearly shown here. Therefore, by monitoring the *RF signal* instead of the *MR signal* there was a gain in SE of around 50%, the SP also improved. An example of the OA simulation and the results of the COB-detector are visible in Figure 6.7, the images also show the pixels used to obtain the different signals.

Table 6.2: Percentage of correct RF pixels detection, PF, and RM pixels erroneously included, PM. MAE comparing the RRs of the CI with the RRs of the *RF signal*, the *RefRF signal*, and the *MR signal*.

Infant	**PF**	**PM**	**MAE (BPM)**		
			RF	**RefRF**	**MR**
1	99.67%	0.00%	0.68	0.67	0.64
2	86.30%	0.15%	1.27	1.26	1.42
4	80.99%	0.35%	3.62	2.51	2.16
6	95.68%	0.13%	0.74	0.71	1.26
10	40.93%	1.21%	4.83	2.98	2.71
11	78.93%	1.19%	1.64	1.30	1.56
14	91.57%	0.00%	2.91	3.14	3.56
15	84.65%	0.12%	2.13	2.00	4.02
17	99.84%	0.00%	2.00	2.11	1.62
Avg.	84.28%	0.35%	2.20	1.85	2.11

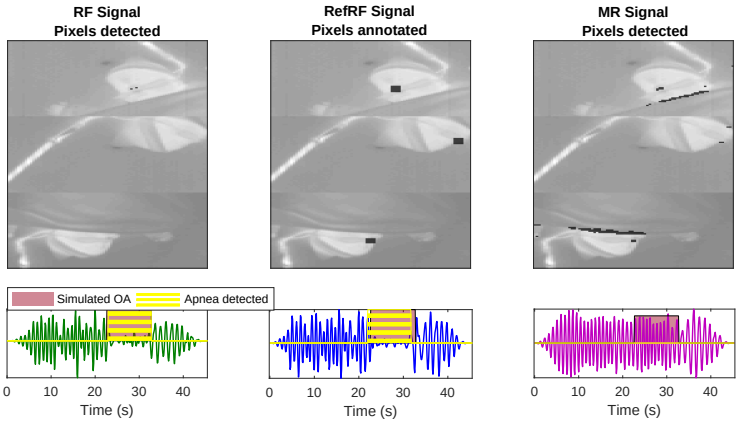


Figure 6.7: Example results of the automatic pixels detection step and of the OA simulation and detection step. The *RefRF signal* shows the simulated OA, the *RF signal* calculated using the automatically detected RF pixels is also able to detect the OA. In the *MR signal*, however, the OA is not visible, being the signal obtained by mixing RM and RF pixels.

6.4 Discussion

The proposed method obtained promising results in the automatic identification of RF pixels in thermal videos, obtaining *RF signals*. Based on our annotation, the time in which the RF was visible in the video segments and where the infants were still, amounts to around 142 minutes split between 11 infants. Therefore,

Table 6.3: Accuracy, sensitivity, and specificity of the OA detection for the *RF signal*, the *RefRF signal*, and the *MR signal*.

Infant	RF			RefRF			MR		
	ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
1	95.49	82.92	96.36	97.74	95.38	98.33	86.09	0.00	100.00
2	88.28	66.78	98.09	97.12	96.76	97.51	76.46	23.83	97.83
4	91.25	61.61	97.32	97.02	95.85	97.15	76.82	31.05	86.42
6	95.42	94.04	96.07	98.43	99.18	98.23	87.09	79.31	87.10
10	89.21	15.96	98.45	97.69	95.68	97.94	82.98	5.73	92.37
11	95.53	74.09	99.86	99.86	95.70	99.65	87.31	6.42	93.09
14	98.16	86.81	100.00	98.31	89.51	99.83	84.11	18.64	94.71
15	95.93	80.82	98.31	98.88	96.41	99.18	76.83	22.37	81.53
17	99.85	97.52	99.89	99.86	96.43	99.98	91.12	22.46	92.50
Avg.	94.35	73.39	98.26	98.32	95.66	98.64	83.20	23.31	91.73

for 4 infants, flow was never visible in the recordings due to the reasons we shall discuss here. The average percentage of still segments with a minimum duration of 30 s that were annotated to have flow visible for the 11 infants is 49% with a maximum of 98% reached by Infant 4.

Different aspects can affect flow visibility in thermal videos. Firstly, the relative position between the infant face and the camera plays an important role in the visibility of RF at the nostrils, and since infants move, the cameras' positions were not always optimal in our study. In addition, the blanket may end up covering the infant's nose/mouth where flow is expected to be visible. The flow visibility annotation was performed by visual inspection of the unprocessed videos, it is possible that flow was present in some recordings but not directly visible due to a low contrast and therefore, neglected in this work. We cannot draw conclusions on whether the low-resolution of our setup had an influence on the flow visibility or on whether the thermal sensitivity of the cameras was insufficient as well. Moreover, other possible factors which make flow detection more complex in infants compared to adults were also mentioned by Abbas et al. in [101], such as the reduced lung volume or the small nasal aperture.

Solutions should aim at maximizing RF visibility in thermal videos. An array of cameras could be used to ensure the visibility of the nostrils area in the videos. Moreover, most of the infants in our dataset are in supine position, which is recommended for infants with a higher PA. However, in NICUs most of the infants in the incubators are in prone position. We expect the prone position to increase the flow visibility on the textiles surrounding the face. Infant 4 was the only infant in prone position in our study and this infant has the highest percentage of flow visibility when the infant is still, as visible in Figure 6.2. Some of the other infants, like 15 or 6, have also flow visible on the textiles even if they are positioned supine, thanks to the head position.

Our automatic RF pixels detection resulted in a percentage of correct RF pixels detected, PF, of 84% as shown in Table 6.2. Based on PM, RM pixels were hardly mixed in, which indicates that the edge removal strategy left a very limited number of RM pixels in the selectable pixels. It should be noted that PM is an estimation of the RM pixels erroneously included, as the RM pixels used for the calculation are the detected MR pixels after removing the annotated RF pixels. This is, therefore, dependent on the MR pixels detected by our previous algorithm, which are selected based on a threshold on the correlation with a core-pixel [180]. PM may be, therefore, underestimated, an annotation of the RM pixels may be needed to accurately estimate PM, the remaining percentage of pixels would then belong to the noise-pixels category. The PF was relatively low for Infant 10, 41%. In some of the recordings with the flow visible at the nostrils, due to a combination of camera position, head position, and temperature gradient between the nose and face, the RF pixels were not correctly identified. Therefore, the removal of the edge caused problems if all RF pixels were located on an edge, which can occur if the flow is visible only at the nostrils and if the relative position between camera

and infant's face causes the nose to be at the edge of the face. Although this last problem could be overcome with a different camera position, further complications are the nose having a lower temperature compared to the face creating additional edges in the image, and the presence of a nasogastric tube which can also create gradients close to the nostrils. The MAE for the *RF signal* was slightly higher than the MAE obtained with a *MR signal*. Compared to the *RefRF signal* MAE, the one for the *RF signal* was significantly higher for infants 4 and 10. The second one is linked with the wrong pixel selection, and similarly, in the case of Infant 4, in some of the segments, noise-pixels were selected instead of RF pixels causing a large MAE. In this case, the flow was also visible on the textiles, however, due to the infant position, the removal of the edges caused the flow region on the textiles to have a more linear shape, the flow was, therefore, not selected. These limitations are further corroborating the need for more cameras and views, which will allow visualizing RF, but also visualizing it away from the image's edges. Our experiments prove that low-cost cameras provide a feasible solution, so adding more cameras should not lead to prohibitive costs, but workflow disturbances should be also considered and minimized.

The proof of concept for the OA detection indicates that there is indeed an advantage in monitoring the *RF signal* compared to the *MR signal*, the sensitivity drastically increased from 23% to 73%, and also the specificity improved as visible in Table 6.3. This result was expected, since the RM pixels were left untouched for the OA simulation. However, if the main source of respiration in the video segments is RF, as is the case in Infant 6, then also the *MR signal* obtained a relatively good sensitivity in the OA detection. The lowest sensitivities for the *RF signal* are linked to the incorrect detection of the RF pixels, indeed infants 10 and 4 resulted in lower sensitivities. This is a proof of concept, and this method should be tested on thermal recordings of real apneas. MA is the most common apnea in infants, characterized by the presence of obstructed inspiratory effort segments as well as central pause segments. Leaving the RM pixels unaltered for the OA simulation is a simplification, as the inspiratory effort of MAs and OAs can present changes in the amplitude and/or frequency compared to the *RM signal* pre-apnea [14].

Moreover, an important discussion point is the presence of ambiguity between the absence of RF, leading to a possible apnea alarm, and the absence of RF visibility. This ambiguity can again be mitigated by increasing the number of camera views, maximizing RF visibility. However, for infants covered with a blanket, this may not be a solution as the blanket could hide the RF. Still, AOP resolves with maturation and, thus, occurs more often in an infant population that is commonly positioned in incubators, i. e. infants with low gestational and postmenstrual age [116]. Infants nursed in incubators are usually not covered by traditional blankets and, thus, the problem is less relevant in this situation. Given an optimal number of views, the ambiguity could still be significant after motion events, which were not studied in this chapter. The method may be combined

with a gross motion detector, similar to the one presented in Chapter 5, and our algorithm for RF pixels detection would need to be reinitialized after a movement of the infant (because of the time dependency introduced in the calculation of the *Covariance Map*). The ambiguity would exist, then, between an apnea occurring after a movement and a position change due to the movement that hides RF completely. Apneas can be preceded by motor activity [181], however, infants are unlikely to completely change posture on their own so as to hide RF from all camera views. If the posture is changed by caregivers, then a protocol could be introduced to make sure the cameras are in a good position to still visualize RF.

Finally, infants in NICUs may need respiratory support to treat severe AOP or to treat other diagnoses. While the nasal cannula may not directly limit the visibility of RF, as in the case of Infant 1 or as shown in [182], the use of a larger interface (nasal mask) will inevitably hide RF at the nostrils. If RF is not visible in the thermal videos due to one of the discussed reasons, RM would anyway be visible in most of the recordings [180], implying that the information available for apnea identification would be similar to the current one, i. e. CI. It should be specified that bradycardia and desaturation need to be also detected in the case of a COB of 10 s to define it an apnea, therefore, for a future complete unobtrusive monitoring of apneas in the NICU these vitals (heartbeat and oxygen saturation) need to be included as well [177].

6.5 Conclusions

The method proposed in this publication was able to detect automatically RF pixels in thermal videos reaching an average percentage of correct pixels detected of 84%. RM pixels were correctly rejected and they were hardly erroneously selected, 0.35%. The MAE was slightly higher on average compared to the one of the *RefRF signal*, 2.20 and 1.85 BPM, respectively. The proof of concept for the OA detection indicates a clear advantage in monitoring a *RF signal* compared to a *MR signal*, the sensitivity increased from 23% to 73%. However, the method should be tested in thermal recordings containing real apneas. RF was annotated to be visible on average in the 49% of the segments in which the infant was still. The number of cameras and their position play an important role in RF visibility in thermal videos, and require further analysis.

CHAPTER 7

Societal Impact

In this thesis, the use of cameras and video algorithms for respiration monitoring and apnea detection has been investigated. Even though healthcare companies are developing camera-based monitoring products, the use of such technology for cardiorespiratory monitoring in clinics is still limited. Studies performed in ideal lab conditions are needed in the initial stage of new technology testing. However, the research in camera-based health monitoring is beyond the initial stage. Therefore, clinical studies are required to move forward and properly assess the usability of cameras in healthcare. Which in turn will allow providing clinicians a clear picture of the situation, avoiding the creation of false expectations. The collaboration between research institutes, hospitals, and industry can deeply contribute to facilitating this process. The work included in this thesis was performed within the Eindhoven MedTech Innovation Centre (e/MTIC) in collaboration with Philips, Máxima Medical Centre, and the Eindhoven University of Technology. This thesis contributes to the analysis of the usability and of the performance of camera technology in the clinic and highlights limitations. The possible long-term impact of this work for the patients and their families, for clinicians, and for industry is discussed in the following sections. Knowledge utilization is also detailed.

Impact for patients and family

Patients would directly benefit from the use of contactless solutions to monitor vital signs as this would reduce the discomfort caused by traditional monitoring methods. In the case of premature infants, the use of contactless solutions helps to reduce the possibility of skin damage. It is clear that the solution developed for the clinical study to monitor respiration using cameras will not substitute all the monitoring needed in critical care or other fields, however, it is a first step towards completely unobtrusive monitoring. For example, in neonatal intensive care units (NICUs) the availability of non-contact respiration monitoring could already allow removing, for some babies, the electrocardiogram (ECG) electrodes. These are the main cause of skin damage due to the adhesive electrodes and are

used to obtain both heart rate (HR) and respiration rate (RR). Considering that HR can be monitored through photoplethysmography (PPG) as well, and RR could be available from camera-based solutions, monitoring ECG would only be necessary for infants with specific diagnoses. Moreover, the use of cameras for home-monitoring has also considerable interest and potential.

Impact for clinicians

The use of cameras and the possibility of monitoring motion may contribute to the reduction of the high false apnea alarm rates currently burdening clinical staff. Motion detection would enable artifact reduction relevant also for other vitals apart from respiration. In addition, gross motion information is considered very relevant for clinical practice and has been linked to the prediction of apneas and discomfort detection. Nurses could also make use of the camera feed to observe the patients from the main nursing station for confirming the necessity of intervention.

Impact for industry

The research conducted proves the possibility of using low-cost thermal cameras for the detection of respiration unobtrusively. Even though the accuracy and the usability of this type of solution still need to be studied further, the possibility to use low-cost thermal solutions is particularly relevant for industry considering the average high cost of traditional thermal cameras.

7.1 Knowledge Utilization

As output of this work three invention disclosures were generated which then resulted in two patent applications filed [P1] and [P2]. Two invention disclosures were combined into a single patent application [P2]. The two patent applications are summarized in the following sections.

7.1.1 Patent application 1

This section provides a summary of the patent application [P1]. Unobtrusive vital signs monitoring for infants is being researched for both hospital and home-care environments. Solutions can involve cameras, radars, or pressure-sensitive films. Some technologies can monitor both respiratory and cardiac activity. Some of these, e. g. continuous-wave radar or single pressure-sensitive film, monitor an entire area. On the other hand, cameras, arrays of pressure-sensitive films, and some radars, e. g. frequency-modulated continuous-wave, require strategies to locate the respiration or heart information in the multidimensional output. Many solutions exploit features to automatically locate cardiac and respiratory activity

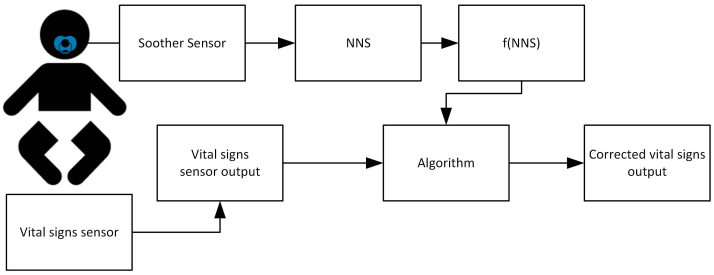


Figure 7.1: Block diagram of proposed solution, adapted from [P1].

signatures in the output and provide information on the RR and the HR. The aforementioned methods and techniques use the implicit assumption that the only periodic signals detectable are the vital signs of interest. Non-nutritive-sucking (NNS) is the motion produced when an infant is moving the soother or pacifier. This motion is periodic and frequencies from 10 to 150 sucks per minute have been reported. Therefore, the presence of NNS can produce errors for the detection of both RR and HR, misidentifying NNS as one of these two vitals. This may cause false bradycardia or tachypnea and it is therefore not desirable. The patent application [P1] proposes the use of a smart soother where sensors can be used to detect the NNS and through algorithm/software solutions the NNS contribution is neglected in the vital signs detection phase. A schematic of the solution is visible in Figure 7.1. Moreover, an example from our data collection when NNS is present in the respiration band is shown in Figure 7.2.

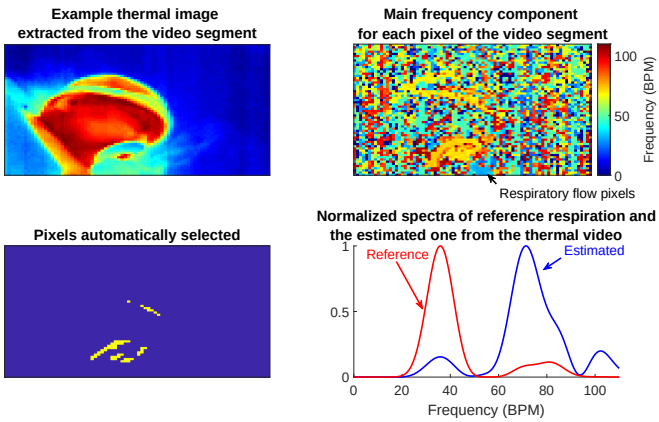
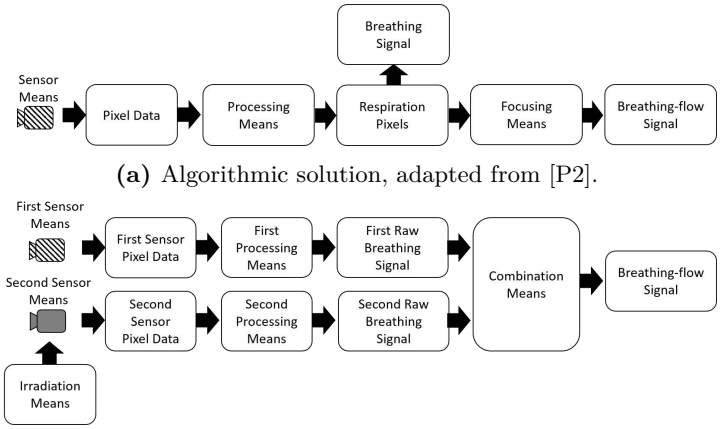


Figure 7.2: Example of results obtained when the NNS frequency is unknown and in the respiration band. The pixels containing NNS are automatically selected instead of the respiratory flow pixels present. Therefore, the wrong RR is detected.

7.1.2 Patent application 2

This section provides a summary of the patent application [P2]. Apnea detection is particularly important in specific environments, e. g. sleep hospitals and NICUs. Considering the different types of apneas, central, obstructive, and mixed, the necessity of monitoring both respiratory motion and flow to differentiate the type is evident. The obtrusiveness of the current monitoring technologies combined with the limitations in apnea detection with obstructive components, e. g. chest impedance (CI), leads to the necessity of unobtrusive solutions able to accurately detect and identify all apnea types. Thermal cameras are the only unobtrusive solution able to monitor both respiratory motion and flow. Recent methods in this field aim at automatic region of interest (ROI) detection to overcome limitations of facial and body landmarks detection. However, these solutions are unable to distinguish contributions coming from respiratory flow and motion in thermal recordings. Therefore, the patent application [P2] proposes algorithmic solutions that exploit characteristics of and differences between respiratory flow and respiratory motion. A possible implementation of this approach has been provided in [J4] and Chapter 6. A general block diagram of this solution is visible in Figure 7.3a. Moreover, considering the differences between non-thermal and thermal videos, i. e. the first visualizes only respiratory motion and the second both flow and motion, it is possible to combine the information from both modalities to automatically obtain a respiratory motion and a respiratory flow signal. If the images coming from the two cameras modalities are aligned, automatic ROI selection strategies could be applied on both and the pixels found in the non-thermal videos can be removed from the thermal video leaving only the respiratory flow pixels. A block diagram is visible in Figure 7.3b.



(a) Algorithmic solution, adapted from [P2].

(b) Combination of thermal and non-thermal imaging.

Figure 7.3: Block diagrams of the solutions proposed.

CHAPTER 8

Conclusions and future research directions

8.1 Conclusions

In critical care settings in hospitals or for sleep monitoring in sleep clinics, vital signs are continuously monitored for diagnostic purposes and/or detecting clinical deterioration and critical events. Respiration is used for the detection of respiratory instability and apneas, i. e. sudden cessations of breathing (COBs). Apneas are typical for two main disorders: sleep-disordered breathing (SDB) and apnea of prematurity (AOP). In neonatal intensive care units (NICUs), respiratory motion is monitored using chest impedance (CI) which is detected through the electrodes used to measure the electrocardiogram (ECG). In sleep clinics, polysomnography (PSG) is used to diagnose and monitor SDB and respiration is detected through respiratory bands and thermistors and/or pressure transducers. The use of sensors requiring contact with the patient causes discomfort and premature infants may also exhibit skin irritation and damage due to the use of adhesive electrodes. Moreover, CI is also unable to detect the occurrence of an apnea if respiratory motion (effort) is still present.

Improved and unobtrusive solutions are required to monitor respiration, detect apneas, and reduce patients' discomfort. Cameras have been researched in this field to monitor several vital signs. Compared to other unobtrusive methods, cameras provide images of the patients useful to obtain contextual information, which can be used for other applications as well e. g. patient observation. Depending on the camera used, different vital signs can be monitored. Focusing on respiration, cameras sensitive in the visible range and near-infrared (NIR) can monitor respiratory motion, whereas thermal cameras in long-wave infrared (LWIR) and mid-wave infrared (MWIR) can monitor both respiratory motion and flow. NIR and thermal cameras are more suited to environments where patients have to be continuously monitored due to the necessity of monitoring vital signs overnight or during poor light conditions. Challenges are still present and require to be addressed before such technology could be employed in the clinic. The objectives of this thesis are: the detection of apneas and COBs

with low latency (Obj. 1), the use of low-cost cameras (Obj. 2), testing in real clinical settings (Obj. 3), automatic respiratory pixels identification (Obj. 4), non-respiratory motion robustness and analysis (Obj. 5), and separating respiratory flow and motion automatically (Obj. 6).

Chapter 2 contributes to Obj. 1 by introducing and describing a method to detect COBs with low latency. The method proposed assumes that a COB can be identified by detecting a significant reduction in the respiration waveform amplitude. The solution is, therefore, based on the comparison of two standard deviations, short-term and long-term. The method has been tested on respiration signals extracted from red-green-blue (RGB) and NIR videos acquired in a neonatal ward. Short COBs were present in the recordings, their occurrence was annotated and they were used to analyze the performance of the algorithm. A method from the literature fit for retrospective use has been applied and the output is used as comparison. The performance of the two methods is comparable, however, our solution can operate on-line. A limitation of the study is the absence of longer apnea events in the recordings.

Chapter 3 contributes to Obj. 2 by investigating the possibility of using thermopile arrays as low-cost alternative to thermal cameras for the detection of respiration. Videos were collected in ideal lab conditions to study the feasibility of the solution. Several distances, respiration rates (RRs), respiratory patterns, and positions have been considered. A method to automatically identify the pixels containing respiration has also been proposed. The study was a proof of concept to demonstrate that thermopile arrays can be used to monitor respiration in ideal conditions, however, the extremely low-resolution of the chosen device limits the possible applications.

Obj. 3 was defined as the introduction of a multi-camera setup in the clinic for the collection of videos. Based on the results of Chapter 3, another low-cost camera solution was identified for the development of the camera setup. Three FLIR Lepton cameras were selected, the resolution is higher than the thermopile array previously used, 60×80 against 8×8 pixels and the sensitivity is improved as well. Three views were considered a minimum requirement to visualize both respiratory flow and motion, i. e. one overview camera for motion, and two looking at the face (left and right) for the flow. A data collection was conducted in the Neonatal Medium Care Unit of the Máxima Medical Centre (MMC) resulting in around 42 hours of thermal videos collected on 15 infants positioned in open bed. The data collected was used in Chapter 4, Chapter 5, and Chapter 6.

Chapter 4 contributes to Obj. 4, and describes an algorithm to merge multiple camera views in a single image plane and the combination of three features able to automatically locate the respiratory pixels. The features designed are independent of facial and body landmarks. This algorithm was compared with a method from literature adapted to be used with our setup. Our method yielded better results and resulted in being able to adapt more easily to window size changes. The methods were tested on part of the thermal videos collected in the Neonatal

Medium Care Unit of the MMC. However, in this study, only the moments in which the infants were not moving were considered.

The work detailed in Chapter 5 directly extends the usability of the algorithm introduced in Chapter 4, and contributes to Obj. 5. The respiration monitoring algorithm (slightly adapted compared to the previously described one) was combined with motion detection and classification. Two main types of motion were identified. Type 1 includes the chest/torso area and type 2 involves the face or limbs. The main assumption behind the extension of the usability of the method was that if non-respiratory motion does not correspond to a type 1 motion, then respiratory motion is still visible and detectable. The performance of the motion detection and the respiration monitoring during the different motions was analyzed. Detecting an RR not only in the moments in which the infants were still (type 2 motion included) allowed to increase the coverage of the method, but the error increased as well. This can be linked to the blanket covering the main source of respiratory motion, i. e. the chest, causing the only respiratory pixels to belong to the areas involved in type 2 motion. The possibility of errors in respiration monitoring caused by the presence of non-nutritive-sucking (NNS) in the recordings is also pointed out. The usability for non-thermal recordings was demonstrated in this work by applying the method to RGB recordings.

In Chapter 6 an algorithm to automatically distinguish respiratory flow from respiratory motion pixels was proposed contributing to Obj. 6. Thermal recordings acquired in the neonatal ward in which respiratory flow was visible were used to develop this method. The algorithm relies on the differences between respiratory motion and flow pixels, and it is based on a combination of features with a bank of Gabor filters. The method proved to be successful in detecting the respiratory flow pixels but limitations are present due to the non-optimal camera views. Moreover, a proof of concept to demonstrate the advantage of monitoring respiratory flow was included. Obstructive apneas (OAs) were simulated in the recordings and the method previously presented in Chapter 2 was used to detect the occurrence of the simulated apneas. Monitoring respiratory flow instead of mixing flow and motion contributions resulted in improving the apnea detectability.

Chapter 7 described the long-term societal impact of this work, highlighting the importance of research for camera-based vital signs monitoring. Two patent applications are summarized and included as knowledge utilization of this thesis. The first proposes a solution to monitor vital signs of an infant unobtrusively and accurately eliminating the possibility of errors caused by NNS, as pointed out in Chapter 5. Therefore, this patent application is contributing to Obj. 5. The second proposes strategies to differentiate between respiratory flow pixels and respiratory motion pixels to obtain an uncontaminated respiratory flow estimation and identify accurately the presence of apneas. One possible solution is the one implemented in Chapter 6, this patent application contributes to Obj. 6.

8.2 Future research

The dataset used in Chapters 4, 5, and 6, was collected in the Neonatal Medium Care Unit of the MMC. While the usability of such a low-cost setup is demonstrated for infants cared for in an open bed, the possibility of introducing the setup in an incubator has to be further analyzed. Incubators are not transparent for thermal radiations and the setup would need to be positioned inside. The small dimension of the cameras would allow this positioning, but proper mounting is required to minimize mechanical risks. Moreover, incubators are heated and humidified environments, therefore a decrease of the thermal contrast should be expected. A clinical study is required to verify how these changes affect the usability of the setup and the algorithms.

The number of cameras and their views should be further analyzed for the different environments to maximize the visibility of both respiratory motion and flow. In particular, the blanket covering the infants in an open bed resulted in hiding the respiratory motion in some of the recordings. This could be a recurrent problem, not only in this environment but also in sleep clinics where subjects can be covered with blankets or sheets. Combining cameras with additional unobtrusive sensors, e. g. radars or pressure-sensitive films, would ensure the availability of respiratory motion from multiple sources. A data fusion could, therefore, mitigate camera-related limitations that result from a hindered view. The visibility of respiratory flow in thermal recordings is also dependent on the camera view and the subject position. An array of thermal cameras is likely required to improve the availability of respiratory flow. Moreover, the usability of the camera solution in the detection of non-simulated apnea events needs to be demonstrated for infants or adults undergoing a sleep study.

It is evident that camera solutions, but also other contactless methods, are unable to provide accurate information on respiratory activity during severe motion. However, this limitation is also present in contact methods commonly used in clinical practice, i. e. CI. Therefore, while it is important to know and understand the limits of a new technology potentially introduced in the hospital, this should not be perceived as a major limitation especially when it is shared with the technology that we are aiming to substitute. Moreover, gross motion monitoring can be considered an additional vital sign since it has been acquiring interest in the clinical field. Motion activity has been indeed linked to applications such as prediction of apneas [167, 183] and discomfort detection [170].

Aiming at remote apnea detection, other vital signs need to be monitored apart from respiration, i. e. oxygen saturation and heart rate (HR). NIR cameras can monitor both in the dark with proper illumination, however studies are still required to overcome limitations due to motion artifacts and skin visibility [177]. Moreover, to achieve continuous monitoring, event detection is required to detect the patient's presence in the bed or occlusions to the camera view, e. g. caused by caregivers' interventions. Machine learning methods have been used with

promising results in this field [79, 184].

Overall, clinical studies are still required to properly analyze the achievable results with cameras compared to contact methods, and to evaluate the accuracy of such solutions. Many other possible applications that could require the monitoring of respiration are not properly discussed in this thesis since the focus is on respiratory disorders characterized by the occurrence of apneas. The use of cameras should be assessed also for applications such as home-care monitoring, baby monitors, or respiratory gating.

Bibliography

[1] J. Tu, K. Inthavong, and G. Ahmadi, "The human respiratory system," in *Computational fluid and particle dynamics in the human respiratory system*, pp. 19–44, Dordrecht: Springer Netherlands, 2013.

[2] L. R. Johnson, "Part III Respiratory physiology," in *Essential medical physiology*, pp. 259–330, Amsterdam: Elsevier, 3rd ed., 2003.

[3] B. H. Culver, "Chapter 4 - Respiratory mechanics," in *Clinical Respiratory Medicine (Third Edition)* (R. K. Albert, S. G. Spiro, and J. R. Jett, eds.), pp. 87–95, Philadelphia: Mosby, 3rd ed., 2008.

[4] I. Wheatley, "Respiratory rate 3: how to take an accurate measurement," *Nursing Times*, vol. 114, no. 7, pp. 21–22, 2018.

[5] G. Yuan, N. A. Drost, and R. A. McIvor, "Respiratory rate and breathing pattern," *McMaster Univ. Med. J*, vol. 10, no. 1, pp. 23–28, 2013.

[6] N. R. Foldvary-Schaefer and T. E. Waters, "Sleep-disordered breathing," *Continuum: Lifelong Learning in Neurology*, vol. 23, no. 4, pp. 1093–1116, 2017.

[7] L. Panossian and J. Daley, "Sleep-disordered breathing," *Continuum: Lifelong Learning in Neurology*, vol. 19, no. 1, pp. 86–103, 2013.

[8] P. J. Strollo and C. W. Atwood, "Chapter 76 - Central sleep apnea and other forms of sleep-disordered breathing," in *Clinical Respiratory Medicine (Third Edition)* (R. K. Albert, S. G. Spiro, and J. R. Jett, eds.), pp. 929–937, Philadelphia: Mosby, 3rd ed., 2008.

[9] G. S. Gilmartin, R. W. Daly, and R. J. Thomas, "Recognition and management of complex sleep-disordered breathing," *Current opinion in pulmonary medicine*, vol. 11, no. 6, pp. 485–493, 2005.

[10] R. Heinzer, S. Vat, P. Marques-Vidal, H. Marti-Soler, *et al.*, "Prevalence of sleep-disordered breathing in the general population: the HypnoLaus study," *The Lancet Respiratory Medicine*, vol. 3, no. 4, pp. 310–318, 2015.

[11] W. Chotinaiwattarakul, L. M. O'Brien, L. Fan, and R. D. Chervin, "Fatigue, tiredness, and lack of energy improve with treatment for OSA," *Journal of Clinical Sleep Medicine*, vol. 5, no. 3, pp. 222–227, 2009.

[12] J. Zhao, F. Gonzalez, and D. Mu, "Apnea of prematurity: from cause to treatment," *European journal of pediatrics*, vol. 170, no. 9, pp. 1097–1105, 2011.

[13] O. Mathew, "Apnea of prematurity: pathogenesis and management strategies," *Journal of Perinatology*, vol. 31, no. 5, pp. 302–310, 2011.

[14] R. J. Martin, "157 - Pathophysiology of apnea of prematurity," in *Fetal and Neonatal Physiology (Fifth Edition)* (R. A. Polin, S. H. Abman, D. H. Rowitch, W. E. Benitz, and W. W. Fox, eds.), pp. 1595–1604, Amsterdam: Elsevier, 5th ed., 2017.

[15] K. J. Barrington and N. N. Finer, "Periodic breathing and apnea in preterm infants," *Pediatric research*, vol. 27, no. 2, pp. 118–121, 1990.

[16] J. M. Di Fiore, R. J. Martin, and E. B. Gauda, "Apnea of prematurity–perfect storm," *Respiratory physiology & neurobiology*, vol. 189, no. 2, pp. 213–222, 2013.

[17] S. Picone, R. Aufieri, and P. Paolillo, "Apnea of prematurity: challenges and solutions," *Research and Reports in Neonatology*, vol. 4, pp. 101–109, 2014.

[18] M. Patel, M. Mohr, D. Lake, J. Delos, *et al.*, "Clinical associations with immature breathing in preterm infants: part 2—periodic breathing," *Pediatric research*, vol. 80, no. 1, pp. 28–34, 2016.

[19] J. V. Rundo and R. Downey, "Chapter 25 - Polysomnography," in *Clinical Neurophysiology: basis and Technical Aspects* (K. H. Levin and P. Chauvel, eds.), vol. 160 of *Handbook of Clinical Neurology*, pp. 381–392, Amsterdam: Elsevier, 2019.

[20] A. Crivello, P. Barsocchi, M. Girolami, and F. Palumbo, "The meaning of sleep quality: a survey of available technologies," *IEEE Access*, vol. 7, pp. 167374–167390, 2019.

[21] O. Bonner, K. Beardsall, N. Crilly, and J. Lasenby, "'There were more wires than him': the potential for wireless patient monitoring in neonatal intensive care," *BMJ innovations*, vol. 3, no. 1, pp. 12–18, 2017.

[22] C. Massaroni, A. Nicolò, D. Lo Presti, M. Sacchetti, *et al.*, "Contact-based methods for measuring respiratory rate," *Sensors*, vol. 19, no. 4, pp. 1–47:908, 2019.

[23] J. M. Di Fiore, "Neonatal cardiorespiratory monitoring techniques," *Seminars in Neonatology*, vol. 9, no. 3, pp. 195–203, 2004.

[24] T. L. Lee-Chiong, "Monitoring respiration during sleep," *Clinics in chest medicine*, vol. 24, no. 2, pp. 297–306, 2003.

[25] F. Q. AL-Khalidi, R. Saatchi, D. Burke, H. Elphick, and S. Tan, "Respiration rate monitoring methods: a review," *Pediatric pulmonology*, vol. 46, no. 6, pp. 523–529, 2011.

[26] R. Budhiraja, J. L. Goodwin, S. Parthasarathy, and S. F. Quan, "Comparison of nasal pressure transducer and thermistor for detection of respiratory events during polysomnography in children," *Sleep*, vol. 28, no. 9, pp. 1117–1121, 2005.

[27] R. Mehra and K. P. Strohl, "Chapter 14 - Evaluation and monitoring of respiratory function," in *Sleep Disorders Medicine (Third Edition)* (S. Chokroverty, ed.), pp. 188–197, Philadelphia: Saunders, 3rd ed., 2009.

[28] E. Schena, C. Massaroni, P. Saccomandi, and S. Cecchini, "Flow measurement in mechanical ventilation: a review," *Medical engineering & physics*, vol. 37, no. 3, pp. 257–264, 2015.

[29] J. Nagler and B. Krauss, "Capnography: a valuable tool for airway management," *Emergency medicine clinics of North America*, vol. 26, no. 4, pp. 881–897, 2008.

[30] K. Zwerneman, "End-tidal carbon dioxide monitoring: a vital sign worth watching.," *Critical care nursing clinics of North America*, vol. 18, no. 2, pp. 217–25, 2006.

[31] A. Aliverti, "Wearable technology: role in respiratory health and disease," *Breathe*, vol. 13, no. 2, pp. e27–e36, 2017.

[32] H. Liu, J. Allen, D. Zheng, and F. Chen, "Recent development of respiratory rate measurement technologies," *Physiological measurement*, vol. 40, no. 7, pp. 1–28:07TR01, 2019.

[33] R. W. van Leuteren, R. E. Bekhuis, C. G. de Waal, F. H. de Jongh, *et al.*, "Diaphragmatic electromyography in preterm infants: the influence of electrode positioning," *Pediatric pulmonology*, vol. 55, no. 2, pp. 354–359, 2020.

[34] S. R. Alty, W. D.-C. Man, J. Moxham, and K. C. Lee, "Denoising of diaphragmatic electromyogram signals for respiratory control and diagnostic purposes," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5560–5563, IEEE, 2008.

[35] P. H. Charlton, T. Bonnici, L. Tarassenko, D. A. Clifton, *et al.*, "An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram," *Physiological measurement*, vol. 37, no. 4, pp. 1–18:610, 2016.

[36] P. H. Charlton, D. A. Birrenkott, T. Bonnici, M. A. Pimentel, *et al.*, "Breathing rate estimation from the electrocardiogram and photoplethysmogram: a review," *IEEE reviews in biomedical engineering*, vol. 11, pp. 2–20, 2017.

[37] V. K. Kapur, D. H. Auckley, S. Chowdhuri, D. C. Kuhlmann, *et al.*, "Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American academy of sleep medicine clinical practice guideline," *Journal of Clinical Sleep Medicine*, vol. 13, no. 3, pp. 479–504, 2017.

[38] A. B. te Pas, "Improving neonatal care with technology," *Frontiers in pediatrics*, vol. 5, pp. 1–3:110, 2017.

[39] S. B. Amin and E. Burnell, "Monitoring apnea of prematurity: validity of nursing documentation and bedside cardiorespiratory monitor," *American journal of perinatology*, vol. 30, no. 8, pp. 643–648, 2013.

[40] C. van Pul, H. van de Mortel, J. van den Bogaart, T. Mohns, and P. Andriessen, "Safe patient monitoring is challenging but still feasible in a neonatal intensive care unit with single family rooms," *Acta Paediatrica*, vol. 104, no. 6, pp. e247–e254, 2015.

[41] K. Lim, C. Eastwood-Sutherland, A. Marshall, T. J. Gale, and P. A. Dargaville, "Limitations of thoracic impedance monitoring for central apnoea detection in preterm infants," *Acta Paediatrica*, 2021. Epub ahead of print.

[42] B. D. Vergales, A. O. Paget-Brown, H. Lee, L. E. Guin, *et al.*, "Accurate automated apnea analysis in preterm infants," *American journal of perinatology*, vol. 31, no. 2, pp. 157–162, 2014.

[43] H. Lee, C. G. Rusin, D. E. Lake, M. T. Clark, *et al.*, "A new algorithm for detecting central apnea in neonates," *Physiological measurement*, vol. 33, no. 1, pp. 1–17, 2011.

[44] N. Kumar, G. Akangire, B. Sullivan, K. Fairchild, and V. Sampath, "Continuous vital sign analysis for predicting and preventing neonatal diseases in the twenty-first century: big data to the forefront," *Pediatric research*, vol. 87, no. 2, pp. 210–220, 2020.

[45] R. Joshi, H. van de Mortel, L. Feijs, P. Andriessen, and C. van Pul, "The heuristics of nurse responsiveness to critical patient monitor and ventilator alarms in a private room neonatal intensive care unit," *PloS one*, vol. 12, no. 10, pp. 1–14:e0184567, 2017.

[46] R. Joshi, C. van Pul, L. Atallah, L. Feijs, *et al.*, "Pattern discovery in critical alarms originating from neonates under intensive care," *Physiological measurement*, vol. 37, no. 4, pp. 564—-579, 2016.

[47] M. A. Mohr, B. D. Vergales, H. Lee, M. T. Clark, *et al.*, "Very long apnea events in preterm infants," *Journal of Applied Physiology*, vol. 118, no. 5, pp. 558–568, 2015.

[48] J. Ede, S. Vollam, J. L. Darbyshire, O. Gibson, *et al.*, "Non-contact vital sign monitoring of patients in an intensive care unit: a human factors analysis of staff expectations," *Applied Ergonomics*, vol. 90, pp. 1–10:103149, 2021.

[49] S. R. Sardesai, M. K. Kornacka, W. Walas, and R. Ramanathan, "Iatrogenic skin injury in the neonatal intensive care unit," *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 24, no. 2, pp. 197–203, 2011.

[50] C. Lund, "Medical adhesives in the NICU," *Newborn and Infant Nursing Reviews*, vol. 14, no. 4, pp. 160–165, 2014.

[51] V. P. Tran, A. A. Al-Jumaily, and S. M. S. Islam, "Doppler radar-based non-contact health monitoring for obstructive sleep apnea diagnosis: a comprehensive review," *Big Data and Cognitive Computing*, vol. 3, no. 1, pp. 1–21:3, 2019.

[52] M. Mercuri, G. Sacco, R. Hornung, P. Zhang, *et al.*, "2-D localization, angular separation and vital signs monitoring using a SISO FMCW radar for smart long-term health monitoring environments," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11065–11077, 2021.

[53] E. Turppa, J. M. Kortelainen, O. Antropov, and T. Kiuru, "Vital sign monitoring using FMCW radar in various sleeping scenarios," *Sensors*, vol. 20, no. 22, pp. 1–19:6505, 2020.

[54] I. Kakouche, A. Maali, M. N. El Korso, A. Mesloub, and M. S. Azzaz, "Fast and cost-effective method for non-contact respiration rate tracking using UWB impulse radar," *Sensors and Actuators A: Physical*, vol. 329, pp. 1–12:112814, 2021.

[55] I. Walterscheid and G. E. Smith, "Respiration and heartbeat monitoring using a distributed pulsed MIMO radar," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3449–3452, IEEE, 2017.

[56] N. Regev and D. Wulich, "Remote sensing of vital signs using an ultra-wide-band radar," *International Journal of Remote Sensing*, vol. 40, no. 17, pp. 6596–6606, 2019.

[57] M. Mercuri, I. Lorato, Y.-H. Liu, F. Wieringa, *et al.*, "Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor," *Nature Electronics*, vol. 2, no. 6, pp. 252–262, 2019.

[58] W. H. Lee, Y. Lee, J. Y. Na, S. H. Kim, *et al.*, "Feasibility of non-contact cardiorespiratory monitoring using impulse-radio ultra-wideband radar in the neonatal intensive care unit," *Plos one*, vol. 15, no. 12, pp. 1–15:e0243939, 2020.

[59] C. E. Goldfine, M. F. T. Oshim, S. P. Carreiro, B. P. Chapman, *et al.*, "Respiratory rate monitoring in clinical environments with a contactless ultra-wideband impulse radar-based sensor system," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, vol. 2020, pp. 3366—-3375, NIH Public Access, 2020.

[60] H. Zhao, H. Hong, D. Miao, Y. Li, *et al.*, "A noncontact breathing disorder recognition system using 2.4-GHz digital-IF doppler radar," *IEEE journal of biomedical and health informatics*, vol. 23, no. 1, pp. 208–217, 2018.

[61] J.-Y. Park, Y. Lee, Y.-W. Choi, R. Heo, *et al.*, "Preclinical evaluation of a noncontact simultaneous monitoring method for respiration and carotid pulsation using impulse-radio ultra-wideband radar," *Scientific reports*, vol. 9, no. 1, pp. 1–12:11892, 2019.

[62] M. Baboli, A. Singh, B. Soll, O. Boric-Lubecke, and V. M. Lubecke, "Wireless sleep apnea detection using continuous wave quadrature Doppler radar," *IEEE Sensors Journal*, vol. 20, no. 1, pp. 538–545, 2019.

[63] W. Li, C. Sun, W. Yuan, W. Gu, *et al.*, "Smart mat system with pressure sensor array for unobtrusive sleep monitoring," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 177–180, IEEE, 2017.

[64] J. J. Liu, M.-C. Huang, W. Xu, X. Zhang, *et al.*, "Breathsens: a continuous on-bed respiratory monitoring system with torso localization using an unobtrusive pressure sensing array," *IEEE journal of biomedical and health informatics*, vol. 19, no. 5, pp. 1682–1688, 2014.

[65] H. Azimi, S. S. Gilakjani, M. Bouchard, S. Bennett, *et al.*, "Breathing signal combining for respiration rate estimation in smart beds," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 303–307, IEEE, 2017.

[66] R. Joshi, B. Bierling, L. Feijs, C. van Pul, and P. Andriessen, "Monitoring the respiratory rate of preterm infants using an ultrathin film sensor embedded in the bedding: a comparative feasibility study," *Physiological measurement*, vol. 40, no. 4, pp. 1–9:045003, 2019.

[67] J. Ranta, T. Aittokoski, M. Tenhunen, and M. Alasaukko-oja, "EMFIT QS heart rate and respiration rate validation," *Biomedical Physics & Engineering Express*, vol. 5, no. 2, pp. 1–8:025016, 2019.

[68] M. Deviaene, I. D. Castro, P. Borzée, A. Patel, *et al.*, "Capacitively-coupled ECG and respiration for the unobtrusive detection of sleep apnea," *Physiological Measurement*, vol. 42, no. 2, pp. 1–14:024001, 2021.

[69] A. Albaba, I. Castro, P. Borzée, B. Buyse, *et al.*, "Automatic quality assessment of capacitively-coupled bioimpedance signals for respiratory activity monitoring," *Biomedical Signal Processing and Control*, vol. 68, pp. 1–11:102775, 2021.

[70] N. Koolen, O. Decroupet, A. Dereymaeker, K. Jansen, *et al.*, "Automated respiration detection from neonatal video data," in *Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM-2015)*, pp. 164–169, 2015.

[71] L. Cattani, D. Alinovi, G. Ferrari, R. Raheli, *et al.*, "A wire-free, non-invasive, low-cost video processing-based approach to neonatal apnoea detection," in *2014 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings*, pp. 67–73, IEEE, 2014.

[72] M. Villarroel, A. Guazzi, J. Jorge, S. Davis, *et al.*, "Continuous non-contact vital sign monitoring in neonatal intensive care unit," *Healthcare technology letters*, vol. 1, no. 3, pp. 87–91, 2014.

[73] J. Jorge, M. Villarroel, S. Chaichulee, A. Guazzi, *et al.*, "Non-contact monitoring of respiration in the neonatal intensive care unit," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 286–293, IEEE, 2017.

[74] C.-Y. Fang, H.-H. Hsieh, and S.-W. Chen, "A vision-based infant respiratory frequency detection system," in *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–8, IEEE, 2015.

[75] D. Alinovi, G. Ferrari, F. Pisani, and R. Raheli, "Respiratory rate monitoring by maximum likelihood video processing," in *2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 172–177, IEEE, 2016.

[76] D. Alinovi, G. Ferrari, F. Pisani, and R. Raheli, "Respiratory rate monitoring by video processing using local motion magnification," in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 1780–1784, IEEE, 2018.

[77] R. Janssen, W. Wang, A. Moço, and G. de Haan, "Video-based respiration monitoring with automatic region of interest detection," *Physiological measurement*, vol. 37, no. 1, pp. 100–114, 2015.

[78] Y. Sun, W. Wang, X. Long, M. Meftah, *et al.*, "Respiration monitoring for premature neonates in NICU," *Applied Sciences*, vol. 9, no. 23, pp. 1–11:5246, 2019.

[79] M. Villarroel, S. Chaichulee, J. Jorge, S. Davis, *et al.*, "Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit," *NPJ digital medicine*, vol. 2, no. 1, pp. 1–18:128, 2019.

[80] C.-W. Wang, A. Hunter, N. Gravill, and S. Matusiewicz, "Unconstrained video monitoring of breathing behavior and application to diagnosis of sleep apnea," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 396–404, 2013.

[81] M. H. Li, A. Yadollahi, and B. Taati, "A non-contact vision-based system for respiratory rate estimation," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2119–2122, IEEE, 2014.

[82] F. Deng, J. Dong, X. Wang, Y. Fang, *et al.*, "Design and implementation of a noncontact sleep monitoring system using infrared cameras and motion sensor," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 7, pp. 1555–1563, 2018.

[83] S. Akbarian, N. M. Ghahjaverestan, A. Yadollahi, and B. Taati, "Distinguishing obstructive versus central apneas in infrared video of sleep using deep learning: validation study," *Journal of Medical Internet Research*, vol. 22, no. 5, pp. 1–14:e17252, 2020.

[84] M. H. Li, A. Yadollahi, and B. Taati, "Noncontact vision-based cardiopulmonary monitoring in different sleeping positions," *IEEE journal of biomedical and health informatics*, vol. 21, no. 5, pp. 1367–1375, 2016.

[85] F. C. Wiegandt, D. Biegger, J. F. Fast, G. Matusiak, *et al.*, "Detection of breathing movements of preterm neonates by recording their abdominal movements with a time-of-flight camera," *Pharmaceutics*, vol. 13, no. 5, pp. 1–13:721, 2021.

[86] L. Cattani, D. Alinovi, G. Ferrari, R. Raheli, *et al.*, "Monitoring infants by automatic video processing: a unified approach to motion analysis," *Computers in biology and Medicine*, vol. 80, pp. 158–165, 2017.

[87] C. Sun, W. Li, C. Chen, Z. Wang, and W. Chen, "An unobtrusive and non-contact method for respiratory measurement with respiratory region detecting algorithm based on depth images," *IEEE Access*, vol. 7, pp. 8300–8315, 2018.

[88] J. Gleichauf, C. Niebler, and A. Koelpin, "Automatic non-contact monitoring of the respiratory rate of neonates using a structured light camera," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 4118–4121, IEEE, 2020.

[89] F. Benetazzo, A. Freddi, A. Monteriù, and S. Longhi, "Respiratory rate detection algorithm based on RGB-D camera: theoretical background and experimental results," *Healthcare technology letters*, vol. 1, no. 3, pp. 81–86, 2014.

[90] A. K. Abbas, K. Heiman, K. Jergus, T. Orlikowsky, and S. Leonhardt, "Neonatal infrared thermography monitoring," in *Neonatal Monitoring Technologies: Design for Integrated Solutions* (W. Chen, S. B. Oetomo, and L. Feijs, eds.), pp. 84–124, Hershey, PA: IGI Global, 2012.

[91] J. Fei and I. Pavlidis, "Thermistor at a distance: unobtrusive measurement of breathing," *IEEE transactions on biomedical engineering*, vol. 57, no. 4, pp. 988–998, 2009.

[92] J. N. Murthy, J. Van Jaarsveld, J. Fei, I. Pavlidis, *et al.*, "Thermal infrared imaging: a novel method to monitor airflow during polysomnography," *Sleep*, vol. 32, no. 11, pp. 1521–1527, 2009.

[93] J. Fei and I. Pavlidis, "Virtual thermistor," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 250–253, IEEE, 2007.

[94] S. L. Bennett, R. Goubran, and F. Knoefel, "Comparison of motion-based analysis to thermal-based analysis of thermal video in the extraction of respiration patterns," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3835–3839, IEEE, 2017.

[95] C. B. Pereira, X. Yu, M. Czaplik, V. Blazek, *et al.*, "Estimation of breathing rate in thermal imaging videos: a pilot study on healthy human subjects," *Journal of clinical monitoring and computing*, vol. 31, no. 6, pp. 1241–1254, 2017.

[96] S. L. Bennett, R. Goubran, and F. Knoefel, "The detection of breathing behavior using Eulerian-enhanced thermal video," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 7474–7477, IEEE, 2015.

[97] C. B. Pereira, X. Yu, M. Czaplik, R. Rossaint, *et al.*, "Remote monitoring of breathing dynamics using infrared thermography," *Biomedical optics express*, vol. 6, no. 11, pp. 4378–4394, 2015.

[98] P. Jakkaew and T. Onoye, "Non-contact respiration monitoring and body movements detection for sleep using thermal imaging," *Sensors*, vol. 20, no. 21, pp. 1–14:6307, 2020.

[99] M. Kopaczka, O. Oezkan, and D. Merhof, "Face tracking and respiratory signal analysis for the detection of sleep apnea in thermal infrared videos with head movement," in *International Conference on Image Analysis and Processing*, pp. 163–170, Springer, 2017.

[100] Y. Cho, S. J. Julier, N. Marquardt, and N. Bianchi-Berthouze, "Robust tracking of respiratory rate in high-dynamic range scenes using mobile thermal imaging," *Biomedical optics express*, vol. 8, no. 10, pp. 4480–4503, 2017.

[101] A. K. Abbas, K. Heimann, K. Jergus, T. Orlikowsky, and S. Leonhardt, "Neonatal non-contact respiratory monitoring based on real-time infrared thermography," *Biomedical engineering online*, vol. 10, no. 1, pp. 1–17:93, 2011.

[102] C. B. Pereira, X. Yu, T. Goos, I. Reiss, *et al.*, "Noncontact monitoring of respiratory rate in newborn infants using thermal imaging," *IEEE transactions on Biomedical Engineering*, vol. 66, no. 4, pp. 1105–1114, 2018.

[103] S. Lyra, L. Mayer, L. Ou, D. Chen, *et al.*, "A deep learning-based camera approach for vital sign monitoring using thermography images for ICU patients," *Sensors*, vol. 21, no. 4, pp. 1–18:1495, 2021.

[104] M.-H. Hu, G.-T. Zhai, D. Li, Y.-Z. Fan, *et al.*, "Synergetic use of thermal and visible imaging techniques for contactless and unobtrusive breathing measurement," *Journal of biomedical optics*, vol. 22, no. 3, pp. 1–11:036006, 2017.

[105] B. Aubakir, B. Nurimbetov, I. Tursynbek, and H. A. Varol, "Vital sign monitoring utilizing Eulerian video magnification and thermography," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3527–3530, IEEE, 2016.

[106] G. Scebba, G. Da Poian, and W. Karlen, "Multispectral video fusion for non-contact monitoring of respiratory rate and apnea," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 1, pp. 350–359, 2020.

[107] W.-H. Chen and H.-P. Ma, "A fall detection system based on infrared array sensors with tracking capability for the elderly at home," in *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)*, pp. 428–434, IEEE, 2015.

[108] G. Sun, T. Saga, T. Shimizu, Y. Hakozaki, and T. Matsui, "Fever screening of seasonal influenza patients using a cost-effective thermopile array with small pixels for close-range thermometry," *International Journal of Infectious Diseases*, vol. 25, pp. 56–58, 2014.

[109] L. J. Goldman, "Nasal airflow and thoracoabdominal motion in children using infrared thermographic video processing," *Pediatric pulmonology*, vol. 47, no. 5, pp. 476–486, 2012.

[110] P. Jagadev and L. I. Giri, "Non-contact monitoring of human respiration using infrared thermography and machine learning," *Infrared Physics & Technology*, vol. 104, pp. 1–13:103117, 2020.

[111] C. B. Pereira, K. Heimann, B. Venema, V. Blazek, *et al.*, "Estimation of respiratory rate from thermal videos of preterm infants," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3818–3821, IEEE, 2017.

[112] P. Chan, G. Wong, T. D. Nguyen, T. Nguyen, *et al.*, "Estimation of respiratory rate using infrared video in an inpatient population: an observational study," *Journal of clinical monitoring and computing*, vol. 34, no. 6, pp. 1275–1284, 2020.

[113] M. van Gastel, S. Stuijk, and G. de Haan, "Robust respiration detection from remote photoplethysmography," *Biomedical optics express*, vol. 7, no. 12, pp. 4941–4957, 2016.

[114] M. Hu, G. Zhai, D. Li, Y. Fan, *et al.*, "Combination of near-infrared and thermal imaging techniques for the remote and simultaneous measurements of breathing and heart rates under sleep situation," *PloS one*, vol. 13, no. 1, pp. 1–14:e0190466, 2018.

[115] E. C. Eichenwald and Committee on fetus and newborn, "Apnea of prematurity," *Pediatrics*, vol. 137, no. 1, pp. 1–7, 2016.

[116] K. Fairchild, M. Mohr, A. Paget-Brown, C. Tabacaru, *et al.*, "Clinical associations of immature breathing in preterm infants: part 1—central apnea," *Pediatric research*, vol. 80, no. 1, pp. 21–27, 2016.

[117] R. T. Brouillette, A. S. Morrow, D. E. Weese-Mayer, and C. E. Hunt, "Comparison of respiratory inductive plethysmography and thoracic impedance for apnea monitoring," *The Journal of pediatrics*, vol. 111, no. 3, pp. 377–383, 1987.

[118] S. A. Pullano, I. Mahbub, M. G. Bianco, S. Shamsir, *et al.*, "Medical devices for pediatric apnea monitoring and therapy: past and new trends," *IEEE reviews in biomedical engineering*, vol. 10, pp. 199–212, 2017.

[119] D. E. Elder, A. J. Campbell, and D. Galletly, "Current definitions for neonatal apnoea: are they evidence based?," *Journal of paediatrics and child health*, vol. 49, no. 9, pp. 388–396, 2013.

[120] V. Cardot, K. Chardon, P. Tourneux, S. Micallef, *et al.*, "Ventilatory response to a hyperoxic test is related to the frequency of short apneic episodes in late preterm neonates," *Pediatric research*, vol. 62, no. 5, pp. 591–596, 2007.

[121] M.-F. Vecchierini, L. Curzi-Dascalova, H. Trang-Pham, J. Bloch, and C. Gaultier, "Patterns of EEG frequency, movement, heart rate, and oxygenation after isolated short apneas in infants," *Pediatric research*, vol. 49, no. 2, pp. 220–226, 2001.

[122] M. Bester, D. Van Den Heever, R. Joshi, and K. Dellimore, "A study of short cessations and temporal dynamics of breathing in preterm infants," in *2018 3rd Biennial South African Biomedical Engineering Conference (SAIBMEC)*, pp. 1–4, IEEE, 2018.

[123] L. A. Aarts, V. Jeanne, J. P. Cleary, C. Lieber, *et al.*, "Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit-—a pilot study," *Early human development*, vol. 89, no. 12, pp. 943–948, 2013.

[124] J. D. Kim, W. H. Lee, Y. Lee, H. J. Lee, *et al.*, "Non-contact respiration monitoring using impulse radio ultrawideband radar in neonates," *Royal Society Open Science*, vol. 6, no. 6, pp. 1–11:190149, 2019.

[125] X. Huang, L. Sun, T. Tian, Z. Huang, and E. Clancy, "Real-time non-contact infant respiratory monitoring using UWB radar," in *2015 IEEE 16th International Conference on Communication Technology (ICCT)*, pp. 493–496, IEEE, 2015.

[126] J. H. Klaessens, M. van den Born, A. van der Veen, J. Sikkens-van de Kraats, *et al.*, "Development of a baby friendly non-contact method for measuring vital signs: first results of clinical measurements in an open incubator at a neonatal intensive care unit," in *Advanced Biomedical and Clinical Diagnostic Systems XII, 89351P*, International Society for Optics and Photonics, 2014.

[127] E. A. Bernal, L. K. Mestha, and E. Shilla, "Non contact monitoring of respiratory function via depth sensing," in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 101–104, IEEE, 2014.

[128] H. Rehouma, R. Noumeir, W. Bouachir, P. Jouvet, and S. Essouri, "3D imaging system for respiratory monitoring in pediatric intensive care

environment," *Computerized Medical Imaging and Graphics*, vol. 70, pp. 17–28, 2018.

[129] A. Bekele, S. Nizami, Y. S. Dosso, C. Aubertin, *et al.*, "Real-time neonatal respiratory rate estimation using a pressure-sensitive mat," in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–5, IEEE, 2018.

[130] S. J. Rhoads, A. L. Green, S. D. Lewis, and L. Rakes, "Challenges of implementation of a web-camera system in the neonatal intensive care unit," *Neonatal Network*, vol. 31, no. 4, pp. 223–228, 2012.

[131] J. Jorge, M. Villarroel, S. Chaichulee, K. McCormick, and L. Tarassenko, "Data fusion for improved camera-based detection of respiration in neonates," in *Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics* (G. L. Coté, ed.), vol. 10501, pp. 215 – 224, International Society for Optics and Photonics, SPIE, 2018.

[132] M. A. Mohr, K. D. Fairchild, M. Patel, R. A. Sinkin, *et al.*, "Quantification of periodic breathing in premature infants," *Physiological measurement*, vol. 36, no. 7, pp. 1415–1427, 2015.

[133] M. T. Clark, B. D. Vergales, A. O. Paget-Brown, T. J. Smoot, *et al.*, "Predictive monitoring for respiratory decompensation leading to urgent unplanned intubation in the neonatal intensive care unit," *Pediatric research*, vol. 73, no. 1, pp. 104–110, 2013.

[134] S. Reuter, C. Moser, and M. Baack, "Respiratory distress in the newborn," *Pediatrics in review*, vol. 35, no. 10, pp. 417–429, 2014.

[135] J. F. Fieselmann, M. S. Hendryx, C. M. Helms, and D. S. Wakefield, "Respiratory rate predicts cardiopulmonary arrest for internal medicine inpatients," *Journal of general internal medicine*, vol. 8, no. 7, pp. 354–360, 1993.

[136] G. F. Lewis, R. G. Gatto, and S. W. Porges, "A novel method for extracting respiration rate and relative tidal volume from infrared thermography," *Psychophysiology*, vol. 48, no. 7, pp. 877–887, 2011.

[137] G. Scebba, L. Tüshaus, and W. Karlen, "Multispectral camera fusion increases robustness of ROI detection for biosignal estimation with nearables in real-world scenarios," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5672–5675, IEEE, 2018.

[138] R. Gade and T. B. Moeslund, "Thermal cameras and applications: a survey," *Machine vision and applications*, vol. 25, no. 1, pp. 245–262, 2014.

[139] J. Fraden, "Physical principles of sensing," in *Handbook of Modern Sensors: Physics, Designs, and Applications*, pp. 69–154, New York: Springer-Verlag, 2016.

[140] G. Mariani and M. Kenyon, "Room-temperature remote sensing: far-infrared imaging based on thermopile technology," in *2015 40th International Conference on Infrared, Millimeter, and Terahertz waves (IRMMW-THz)*, pp. 1–2, IEEE, 2015.

[141] J. Tanaka, M. Shiozaki, F. Aita, T. Seki, and M. Oba, "Thermopile infrared array sensor for human detector application," in *2014 IEEE 27th International Conference on Micro Electro Mechanical Systems (MEMS)*, pp. 1213–1216, IEEE, 2014.

[142] N. Deckers, M. Yildirim, and R. Reulke, "Sensor fusion-based learning for the improvement of person segmentation by means of a low-resolution thermal infrared array sensor," in *Proceedings of the 2017 International Conference on Computer Graphics and Digital Image Processing*, pp. 1–6:9, ACM, 2017.

[143] A. D. Shetty, B. Shubha, and K. Suryanarayana, "Detection and tracking of a human using the infrared thermopile array sensor—"Grid-EYE"," in *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, pp. 1490–1495, IEEE, 2017.

[144] F.-F. Lee, F. Chen, and J. Liu, "Infrared thermal imaging system on a mobile phone," *Sensors*, vol. 15, no. 5, pp. 10166–10179, 2015.

[145] A. K. Abbas, K. Heiman, T. Orlikowsky, and S. Leonhardt, "Non-contact respiratory monitoring based on real-time IR-thermography," in *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*, pp. 1306–1309, Springer, 2009.

[146] M. Troost, *Presence detection and activity recognition using low-resolution passive IR sensors.* MSc. Thesis, Technische Universiteit Eindhoven, 2013.

[147] P. Höppe, "Temperatures of expired air under varying climatic conditions," *International journal of biometeorology*, vol. 25, no. 2, pp. 127–132, 1981.

[148] I. Smith, J. Mackay, N. Fahrid, and D. Krucheck, "Respiratory rate measurement: a comparison of methods," *British Journal of Healthcare Assistants*, vol. 5, no. 1, pp. 18–23, 2011.

[149] S. G. Fleming and L. Tarassenko, "A comparison of signal processing techniques for the extraction of breathing rate from the photoplethysmogram," *Int J Biol Med Sci*, vol. 2, no. 4, pp. 232–6, 2007.

[150] S. M. Kay and S. L. Marple, "Spectrum analysis—a modern perspective," *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1380–1419, 1981.

[151] J. Pardey, S. Roberts, and L. Tarassenko, "A review of parametric modelling techniques for EEG analysis," *Medical engineering & physics*, vol. 18, no. 1, pp. 2–11, 1996.

[152] J. L. Salinet, N. Masca, P. J. Stafford, G. A. Ng, and F. S. Schlindwein, "Three-dimensional dominant frequency mapping using autoregressive spectral analysis of atrial electrograms of patients in persistent atrial fibrillation," *Biomedical engineering online*, vol. 15, no. 1, pp. 1–15:28, 2016.

[153] S. Aydin, "Determination of autoregressive model orders for seizure detection," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 18, no. 1, pp. 23–30, 2010.

[154] S. Chaichulee, M. Villarroel, J. Jorge, C. Arteta, *et al.*, "Localised photoplethysmography imaging for heart rate estimation of pre-term infants in the clinic," in *Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics*, vol. 10501, pp. 146 – 159:105010R, International Society for Optics and Photonics, 2018.

[155] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.

[156] N. Rutter, "Applied physiology: the newborn skin," *Current Paediatrics*, vol. 13, no. 3, pp. 226–230, 2003.

[157] C. Battersby, S. Michaelides, M. Upton, and J. M. Rennie, "Term admissions to neonatal units in England: a role for transitional care? a retrospective cohort study," *BMJ open*, vol. 7, no. 5, pp. 1–7:e016050, 2017.

[158] S. Nizami, A. Bekele, M. Hozayen, K. Greenwood, *et al.*, "Comparing time and frequency domain estimation of neonatal respiratory rate using pressure-sensitive mats," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 239–244, IEEE, 2017.

[159] M. Usman, R. Evans, R. Saatchi, R. Kingshott, and H. Elphick, "Non-invasive respiration monitoring by thermal imaging to detect sleep apnoea," in *The 32nd International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management*, September 2019. http://shura.shu.ac.uk/id/eprint/24964 (Unpublished).

[160] I. Lorato, T. Bakkes, S. Stuijk, M. Meftah, and G. De Haan, "Unobtrusive respiratory flow monitoring using a thermopile array: a feasibility study," *Applied Sciences*, vol. 9, no. 12, pp. 1–15:2449, 2019.

[161] M. M. Baharestani, "An overview of neonatal and pediatric wound care knowledge and considerations," *Ostomy/wound management*, vol. 53, no. 6, pp. 34–55, 2007.

[162] N. Bu, N. Ueno, and O. Fukuda, "Monitoring of respiration and heartbeat during sleep using a flexible piezoelectric film sensor and empirical mode decomposition," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1362–1366, IEEE, 2007.

[163] D. Shao, C. Liu, and F. Tsow, "Noncontact physiological measurement using a camera: a technical review and future directions," *ACS sensors*, vol. 6, no. 2, pp. 321–334, 2020.

[164] C. Massaroni, A. Nicolò, M. Sacchetti, and E. Schena, "Contactless methods for measuring respiratory rate: a review," *IEEE Sensors Journal*, vol. 21, no. 11, pp. 12821–12839, 2020.

[165] E. Mercuri, M. C. Pera, and C. Brogna, "Chapter 21 - Neonatal hypotonia and neuromuscular conditions," in *Neonatal Neurology* (L. S. de Vries and H. C. Glass, eds.), vol. 162 of *Handbook of Clinical Neurology*, pp. 435–448, Amsterdam: Elsevier, 2019.

[166] E. M. Mizrahi and R. R. Clancy, "Neonatal seizures: early-onset seizure syndromes and their consequences for development," *Mental retardation and developmental disabilities research reviews*, vol. 6, no. 4, pp. 229–241, 2000.

[167] K. Lim, H. Jiang, A. P. Marshall, B. Salmon, *et al.*, "Predicting apnoeic events in preterm infants," *Frontiers in Pediatrics*, vol. 8, pp. 1–7:570, 2020.

[168] R. Joshi, D. Kommers, L. Oosterwijk, L. Feijs, *et al.*, "Predicting neonatal sepsis using features of heart rate variability, respiratory characteristics, and ECG-derived estimates of infant motion," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 3, pp. 681–692, 2019.

[169] I. Lorato, S. Stuijk, M. Meftah, D. Kommers, *et al.*, "Multi-camera infrared thermography for infant respiration monitoring," *Biomedical Optics Express*, vol. 11, no. 9, pp. 4848–4861, 2020.

[170] Y. Sun, D. Kommers, W. Wang, R. Joshi, *et al.*, "Automatic and continuous discomfort detection for premature infants in a NICU using video-based motion analysis," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5995–5999, IEEE, 2019.

[171] M. Hafström, C. Lundquist, K. Lindecrantz, K. Larsson, and I. Kjellmer, "Recording non-nutritive sucking in the neonate. Description of an automatized system for analysis," *Acta Paediatrica*, vol. 86, no. 1, pp. 82–90, 1997.

[172] R. Pineda, K. Dewey, A. Jacobsen, and J. Smith, "Non-nutritive sucking in the preterm infant," *American journal of perinatology*, vol. 36, no. 3, pp. 268–276, 2019.

[173] I. Lorato, S. Stuijk, M. Meftah, W. Verkruijsse, and G. de Haan, "Camera-based on-line short cessation of breathing detection," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 1656–1663, IEEE, 2019.

[174] P. C. Loughlin, F. Sebat, and J. G. Kellett, "Respiratory rate: the forgotten vital sign—make it count!," *Joint Commission journal on quality and patient safety*, vol. 44, no. 8, pp. 494–499, 2018.

[175] G. Erickson, N. R. Dobson, and C. E. Hunt, "Immature control of breathing and apnea of prematurity: the known and unknown," *Journal of Perinatology*, pp. 1–13, 2021. Epub ahead of print.

[176] S. Schellenberger, K. Shi, F. Michler, F. Lurz, *et al.*, "Continuous in-bed monitoring of vital signs using a multi radar setup for freely moving patients," *Sensors*, vol. 20, no. 20, pp. 1–14:5827, 2020.

[177] M. van Gastel, S. Stuijk, S. Overeem, J. P. van Dijk, *et al.*, "Camera-based vital signs monitoring during sleep–a proof of concept study," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1409–1418, 2020.

[178] M. Martinez and R. Stiefelhagen, "Breathing rate monitoring during sleep from a depth camera under real-life conditions," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1168–1176, IEEE, 2017.

[179] W. Imano, K. Kameyama, M. Hollingdal, J. Refsgaard, *et al.*, "Non-contact respiratory measurement using a depth camera for elderly people," *Sensors*, vol. 20, no. 23, pp. 1–12:6901, 2020.

[180] I. Lorato, S. Stuijk, M. Meftah, D. Kommers, *et al.*, "Towards continuous camera-based respiration monitoring in infants," *Sensors*, vol. 21, no. 7, pp. 1–18:2268, 2021.

[181] C. F. Poets, "Apnea of prematurity," in *Sleep Disordered Breathing in Children: a Comprehensive Clinical Guide to Evaluation and Treatment* (L. Kheirandish-Gozal and D. Gozal, eds.), pp. 263–278, Totowa, NJ: Humana Press, 2012.

[182] N. Hochhausen, C. Barbosa Pereira, S. Leonhardt, R. Rossaint, and M. Czaplik, "Estimating respiratory rate in post-anesthesia care unit patients using infrared thermography: an observational study," *Sensors*, vol. 18, no. 5, pp. 1–12:1618, 2018.

[183] I. Zuzarte, D. Paydarfar, and D. Sternad, "Effect of spontaneous movement on respiration in preterm infants," *Experimental Physiology*, vol. 106, no. 5, pp. 1285–1302, 2021.

[184] R. Weber, S. Cabon, A. Simon, F. Porée, and G. Carrault, "Preterm newborn presence detection in incubator and open bed using deep transfer learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1419–1428, 2021.

Acknowledgements

During these years, I sometimes felt that the PhD journey resembled a quote from the children's novel The Bad Beginning of Lemony Snicket, *"At times the world may seem an unfriendly and sinister place, but believe that there is much more good in it than bad. All you have to do is look hard enough. and what might seem to be a series of unfortunate events may in fact be the first steps of a journey."*, just substitute "world" with "PhD". Both overcoming these unfortunate events and achieving publishable and novel results have been a shared effort and therefore, I have many people to thank.

Firstly, I would like to thank my promotor, Gerard de Haan, for allowing me to work on this project. In my first years, I naively thought that at the end of meetings I would have the answers to my questions, but they always ended with more questions (and also some answers, of course). Therefore, thank you Gerard for asking the right questions. My gratitude goes also to my 1st copromotor, Sander Stuijk and to his sharp mind. I remember asking you if all PhD candidates had to face so many problems and you said that if that were true you would not have had any time left. Thank you for always being available and for helping me solve the many challenges I had to face in these years regarding research, technical, and administrative problems. My appreciation goes to my 2nd copromotor Carola van Pul for representing the much needed connection between the hospital and university worlds. Carola thank you for your invaluable biomedical knowledge, for your unlimited availability, and for sharing the Dutch culture with us. I honestly do not know where you find the time and the energy!

Furthermore, I would like to thank Mohammed Meftah, my supervisor at Philips. Mohammed, I learned a lot from your experience with clinical studies, so much that I started wondering "what would Mohammed say?" when planning a data collection. I can only hope we will have the chance to work together again. I am grateful for the support I received from Peter Andriessen, my supervisor at MMC, I got here also thanks to your contagious enthusiasm and endless curiosity. I thank also Ward Cottaar, Xi Long, and all the other members of the ALARM team for their collaboration and feedback. My gratitude extends to the committee members who agreed to be part of the defense committee and gave me feedback on the manuscript.

Moreover, I would like to thank the people that helped me in MMC. To the senior nurses in the neonatal ward, Astrid Osagiator and Heidi van de Mortel, thank you for your kindness and availability. I thank all the nurses of the Medium Care Unit for their patience and for helping with my data collection even in these difficult

times. I further extend my gratitude to the parents of the babies who agreed to participate in the study.

Additionally, I would like to thank Marja de Mol - Regels and Wendy Verhoeven - Houben for their help and for the organization of the events in the ES group. A thank is owed also to Martijn Koedam for helping me with my setup, ordering all the materials necessary, and being always available for technical problems. Thanks to Mark Wijtvliet for his help in soldering back one of the cameras' connectors and for printing the cameras' housings. I thank also the students that I helped supervise for their contributions, Tom, Imke, Ineke, and Evelien. Furthermore, thanks to Benoit Balmaekers, Marek Bartula, Mark van Gastel, Marko de Jager, Rik Janssen, Shakith Fernando, Wenjin Wang, Wim Verkruijsse, and all the other people from Philips Research with whom I collaborated.

A special thanks to colleagues and friends. My gratitude goes to my master thesis supervisor and now friend Marco Mercuri, I probably wouldn't have been here if our paths did not randomly cross, so thank you! Gabriele, Deedee, Laura, and Zheng thanks for the lunches and for the great collaboration. Many thanks to Anouk, Luc, and Savvas for the Friday dinners and drinks. It became a much needed tradition to forget the challenges of the week. Kamlesh, Paul, Sayandip, and Shima thank you for the lunches, the nights out, and the travels. A big thank you to my office mates Amr, Gagandeep, Hamideh, and Mojtaba for sharing the best office on the 4th floor. Joan, Nasim, and Sayra, I hope you will enjoy it as much as we did! Ramon thank you for being a breath of fresh air, and Barry thank you for being always present and active in the group. Furthermore, thanks to Berk, Martin, Greice, Sherif, Emad, David, Raquel, and all the other colleagues of the ES group for the interesting talks and environment.

I thank my family for being a constant presence in my life. To my sister, Giuliana thank you for being my "sister/friend" and for reading all my long messages. To my twin brother, Luigi thank you for being our only family in The Netherlands for a while. A thank is owed to my parents for their support and, more importantly, for the packages of food that kept us close to Italy in these five years. Finally, Ale, we had the luck of sharing the PhD experience and understanding each other's difficulties and achievements. Thanks for making Eindhoven my home away from home.

Curriculum Vitae

Ilde Lorato was born in Potenza, Italy in 1992. She received the bachelor's degree in Biomedical Engineering from the University of Pisa, Pisa, Italy in 2014. During her master's studies, she moved to Eindhoven to carry out the work for the master's thesis at Holst Centre- Imec NL focusing on the use of radars for unobtrusive vital signs monitoring. She graduated cum laude for the master's degree in Biomedical Engineering at the University of Pisa, Pisa, Italy in 2017. Afterwards, she started her PhD in the Electronic Systems group at the Eindhoven University of Technology (TU/e). Her research is part of the Alarm-Limiting AlgoRithm-based Monitoring (ALARM) project funded by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek with grant number 15345. Máxima Medical Centre and Philips Research were the clinical and industrial partners, respectively. Her research interests include medical technology, unobtrusive vital signs monitoring, and image/signal processing.

List of publications

Journals

[J1] I. Lorato, T. Bakkes, S. Stuijk, M. Meftah, and G. de Haan, "Unobtrusive respiratory flow monitoring using a thermopile array: a feasibility study," Applied Sciences, vol. 9, no. 12, pp. 1–15:2449, 2019.

[J2] I. Lorato, S. Stuijk, M. Meftah, D. Kommers, P. Andriessen, C. van Pul, and G. de Haan, "Multi-camera infrared thermography for infant respiration monitoring," Biomedical Optics Express, vol. 11, no. 9, pp. 4848–4861, 2020.

[J3] I. Lorato, S. Stuijk, M. Meftah, D. Kommers, P. Andriessen, C. van Pul, and G. de Haan, "Towards continuous camera-based respiration monitoring in infants," Sensors, vol. 21, no. 7, pp. 1–18:2268, 2021.

[J4] I. Lorato, S. Stuijk, M. Meftah, D. Kommers, P. Andriessen, C. van Pul, and G. de Haan, "Automatic separation of respiratory flow from motion in thermal videos for infant apnea detection," Sensors, vol. 21, no. 18, pp. 1–16:6306, 2021.

Conference proceedings

[C1] I. Lorato, S. Stuijk, M. Meftah, W. Verkruijsse, and G. de Haan, "Camera-based on-line short cessation of breathing detection," in 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp. 1656–1663, IEEE, 2019.

Patent applications

[P1] I. Lorato, "A method and system for monitoring vital signs of an infant." First filing, Jan. 2021.

[P2] G. de Haan and I. Lorato, "Device, system and method for determining respiratory information of a subject." First filing, July 2021.

Publications and patents not covered in this thesis

[NC1] Z. Peng, I. Lorato, X. Long, R.-H. Liang, D. Kommers, P. Andriessen, W. Cottaar, S. Stuijk, and C. van Pul, "Body motion detection in neonates based on motion artifacts in physiological signals from a clinical patient monitor," accepted for 2021 IEEE/The Engineering in Medicine and Biology Conference (EMBC).

[NC2] M. Mercuri and I. Lorato, "Method and a system for localization and monitoring of living being targets." US Patent 11,009,599, May 2021.

[NC3] I. Lorato, "Method and a system for detecting a vital sign of a subject." US Patent 10,966,663, Apr. 2021.

[NC4] M. Mercuri, I. Lorato, Y.-H. Liu, F. Wieringa, C. Van Hoof, and T. Torfs, "Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor," Nature Electronics, vol. 2, no. 6, pp. 252–262, 2019.

[NC5] M. Mercuri, Y.-H. Liu, I. Lorato, T. Torfs, F. Wieringa, A. Bourdoux, and C. Van Hoof, "A direct phase-tracking Doppler radar using wavelet independent component analysis for non-contact respiratory and heart rate monitoring," IEEE transactions on biomedical circuits and systems, vol. 12, no. 3, pp. 632–643, 2018.

[NC6] I. D. Castro, M. Mercuri, T. Torfs, I. Lorato, R. Puers, and C. Van Hoof, "Sensor fusion of capacitively coupled ECG and continuous-wave Doppler radar for improved unobtrusive heart rate measurements," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 8, no. 2, pp. 316–328, 2018.

[NC7] M. Mercuri, Y.-H. Liu, I. Lorato, T. Torfs, A. Bourdoux, and C. Van Hoof, "Frequency-tracking CW Doppler radar solving small-angle approximation and null point issues in non-contact vital signs monitoring," IEEE transactions on biomedical circuits and systems, vol. 11, no. 3, pp. 671–680, 2017.