UNIVERSITÀ DEGLI STUDI DI PARMA
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

*Dottorato di Ricerca in Tecnologie dell'Informazione*

*XXIX Ciclo*

Davide Alinovi

# VIDEO PROCESSING FOR REMOTE RESPIRATION MONITORING

Dissertazione presentata per il conseguimento
del titolo di Dottore di Ricerca

Febbraio 2017

# UNIVERSITÀ DEGLI STUDI DI PARMA

*Dottorato di Ricerca in Tecnologie dell'Informazione*

*XXIX Ciclo*

# VIDEO PROCESSING FOR REMOTE RESPIRATION MONITORING

Relatore:
*Chiar.mo Prof. Riccardo Raheli*

Coordinatore:
*Chiar.mo Prof. Marco Locatelli*

Dottorando: *Davide Alinovi*

Febbraio 2017

*To my family.*

# Abstract

Monitoring of vital signs is a key tool in medical diagnostics to asses the onset and the evolution of several diseases. Among fundamental vital parameters, such as the hearth rate, blood pressure and body temperature, the Respiratory Rate (RR) plays an important role. For this reason, respiration needs to be carefully monitored in order to detect potential signs or events indicating possible changes of health conditions.

Monitoring of the respiration is generally carried out in hospital and clinical environments by the use of expensive devices with several sensors connected to the patient's body. A new research trend, in order to reduce healthcare service costs and make monitoring of vital signs more comfortable, is the development of low-cost systems which may allow remote and contactless monitoring; in such a context, an appealing method is to rely on video processing-based solutions.

In this dissertation, novel techniques for the visualization and analysis of respiration by remote video monitoring, based on the study of breathing-related movements, are proposed. Due to the modest extent of movements related to respiration in both infants and adults, specific algorithms in order to efficiently detect breathing are needed. For this reason, motion-related variations in video signals are exploited with various algorithms to identify respiration of the monitored patient and simultaneously estimate the RR over time.

In particular, three algorithms are proposed and analyzed. In one solution, video signals are first processed for motion magnification and then analyzed for the monitoring of respiratory movements. In a second solution, the subtle motion amplification is integrated with the video-processing algorithm for the analysis of movements, with the aim to simplify the processing and improve the performance. In a third solution, a direct Maximum Likelihood (ML)-based video processing algorithm for periodicity analysis is investigated.

An important feature of the proposed algorithms is the possibility to detect temporary absence of breathing movements, which may be the indicator of serious diseases, like apneas. Such events can be also caused by life-threatening diseases, that need timely treatments; moreover, in newborns, there may be a risk of Sudden Infant Death Syndrome (SIDS).

A significant problem which makes the design and optimization of video processing-based monitoring systems quite critical, is the lack of large video databases, associated with clinical data, obtained from real patients with respiratory disorders. Hence, a Continuous-Time Markov Chain (CTMC) statistical model of breathing patterns, including the possibility to describe respiratory pauses and apnea events, is proposed. The model, tuned and driven by real data extracted from monitored patients, is able to describe realistic breathing patterns. Then, two suitable simulators, software- and hardware-based, have been developed, demonstrating that the proposed CTMC-based statistical model can be strategic to devise simulators useful to test and design novel and effective video processing-based monitoring systems.

Finally, performance evaluation of the proposed video processing-based algorithms is performed, by the use of the previously proposed simulators (hardware- and software-based), as well as real cases.

# Contents

# List of Acronyms

| | | |
|---|---|---|
| **AUC** | Area Under Curve | 109 |
| **CCHS** | Congenital Central Hypoventilation Syndrome | 9, 75, 106 |
| **CTMC** | Continuous-Time Markov Chain | 3, 75, 99, 114 |
| **DFT** | Discrete Fourier Transform | 32, 85 |
| **DoF** | Difference of Frames | 29 |
| **DOR** | Diagnostic Odds Ratio | 109 |
| **ECG** | ElectroCardioGram | 15 |
| **EEG** | ElectroEncephaloGram | 14, 41, 90 |
| **EMG** | ElectroMyoGram | 15 |
| **EOG** | ElectroOculoGram | 15 |
| **EVM** | Eulerian Video Magnification | 35, 113 |
| **FBG** | Fiber Bragg Gratings | 17 |
| **FIR** | Finite Impulse Response | 62 |
| **HR** | Heart Rate | 7, 15, 115 |
| **i.i.d.** | independent and identically distributed | 32, 77 |

| | | |
|---|---|---|
| **IIR** | Infinite Impulse Response | 37 |
| **KL** | Kullback-Leibler | 93 |
| **ML** | Maximum Likelihood | 3, 32, 85, 113 |
| **NICU** | Neonatal Intensive Care Unit | 16, 44, 90, 105 |
| **OSAS** | Obstructive Sleep Apnea Syndrome | 8 |
| **PMF** | Probability Mass Function | 92 |
| **RGB** | Red, Green and Blue | 23, 28 |
| **RMS** | Root Mean Squared | 56, 111 |
| **RMSE** | Root Mean Squared Error | 56, 111 |
| **ROC** | Receiver Operating Characteristic | 109 |
| **ROI** | Regions of Interest | 58, 114 |
| **RR** | Respiratory Rate | 1, 7, 13, 27, 75, 98, 113 |
| **SIDS** | Sudden Infant Death Syndrome | 8 |
| **TD** | Tolerance Delay | 43, 107 |

# Introduction

The main focus of this work is on efficient monitoring of vital signs, intended as the measurement, by appropriate devices, of human bodily functions. Main vital signs monitored by healthcare providers include, e.g., heart rate, blood pressure, rate of breathing and body temperature. Monitoring of these vital signs has the task to help clinicians in assessing the health conditions of a patient and check for the possible outbreak and the evolution of physical disorders.

One of the major physiological tasks that need to be carefully monitored in patients is the respiration. Among parameters used to control the efficiency of the respiration system, the Respiratory Rate (RR) plays a fundamental role. In fact, anomalous breathing rates can be significant indicators of a hidden pathological condition. Specifically, they could be useful to predict life-threatening events or reveal hazardous absence of respiration, as an apnea event. For this reason, the RR needs to be continuously monitored in order to recognize potential evidence of possible changes of a patient's health conditions.

Nowadays, the gold-standard methods for respiration monitoring mainly employed in clinical environments are based on devices that require a direct connection to the patient's body. These devices cannot always be used in home environments or applied for telemedicine or in eHealth contexts. Furthermore, direct connection of the sensors to the patient, makes these systems quite uncomfortable, particularly if applied on infants.

In recent times, the research and development of portable, comfortable and even non-contact equipments for the monitoring of respiration and other vital signs experienced a dramatic growth, also thanks to the miniaturization and drop of price of the sensors involved. One of the most appealing methods employed for monitoring tasks and to create possible automatic diagnostic tools is based on image and video processing.

This thesis is focused on the development and analysis of novel video processing-based algorithms for the measurement of the RR and the detection of specific events related to breathing disorders, such as apneas. In particular, innovative techniques to evaluate the performance of these monitoring systems are proposed.

The dissertation is organized as follows. Before describing the developed algorithms and their performance, in Chapter 1 the breathing mechanisms in humans and related diseases are introduced. In particular, respiratory disorders in newborns are described, focusing on apneas and congenital breathing disorders, which will be considered later for the development of video processing-based algorithms.

The thesis is then divided in two parts. The first part of this work investigates respiration monitoring and the analysis of breathing movements for the estimation of the RR and the detection of apneas.

In Chapter 2, the standard methods employed by medical staff in hospitals for monitoring of respiration and other physiological parameters are introduced. Then, starting with an overview on alternative contact-based and wearable systems, the focus is moved to innovative non-invasive and contactless devices. Within this category, a specific section is dedicated to video processing based systems.

In Chapter 3, novel techniques for the visualization and analysis of respiration by remote video monitoring, based on the study of breathing-related movements, are proposed. Since motion related to respiration is difficult to detect as temporal variation in video signals, especially in newborns, these small changes are exploited in order to identify the breathing effort and estimate the RR. For this purpose, three algorithms are proposed and evaluated. In the first solution, video signals are pre-processed for motion magnification to enhance breathing movements; these are then analyzed for the evaluation of the possible occurrence of apneas in the monitored patient. In a second solution, the motion amplification technique is integrated with the video-processing algorithm for the analysis of movements, with the aim to simplify the processing and improve the performance. In a third solution, a direct Maximum Likelihood (ML) video processing algorithm for periodicity analysis is investigated, comparing the RR estimation performance with previous solutions.

The second part of this dissertation deals with a problem which makes the design and optimization of video processing-based monitoring systems critical: the lack of large and accurate databases of video streams recorded from real patients.

In Chapter 4, a Continuous-Time Markov Chain (CTMC)-based statistical model of breathing patterns, including the possibility to describe respiratory pauses and apnea events, is proposed. The model, tuned and driven by real data extracted from patients monitored by standard clinical devices, can describe realistic breathing patterns.

Two kinds of “simulators of respiration” drawn on the statistical model previously introduced, have been developed and carefully described in Chapter 5. Finally, performance evaluation of the video processing-based algorithm for apnea detection proposed in Chapter 3 is carried on by the use of the two proposed simulators.

### Standard Protocol Approvals, Registrations, and Patient Consents

In accordance with current practice at our Institution, an informed consent form was signed by a parent of each newborn patient, and the aforementioned document was stored in the patients’ hospital chart. Analysis and use of biomedical signals and video recordings was approved by the Ethical Local Committee.

# Chapter 1

# Respiration in Humans

## 1.1 Introduction

Respiration is an important physiological task in living organisms and plays a fundamental role in human body functions.

In this chapter, the fundamentals of the physiology of the respiratory system and related diseases are introduced. The chapter is organized as follows. In Section 1.2, the mechanisms involved in human respiration are introduced; then, breathing dysfunctions in newborns, especially the ones of interest in this work, are described in Section 1.3.

## 1.2 Respiration System and Mechanisms

The primary function of the respiratory system is gas exchange. Oxygen ($O_2$), which is needed for the functionality of cells, is transferred from the exterior into the blood, while carbon dioxide ($CO_2$), a "waste" product of cellular function, is expelled into the outside air [1]. This process results in air containing oxygen being inhaled into the lungs, where gas exchange occurs across the alveolar-capillary membrane. Then, $CO_2$ is excreted in the air and released through the mouth or nose. The global process including inhalation and ex-

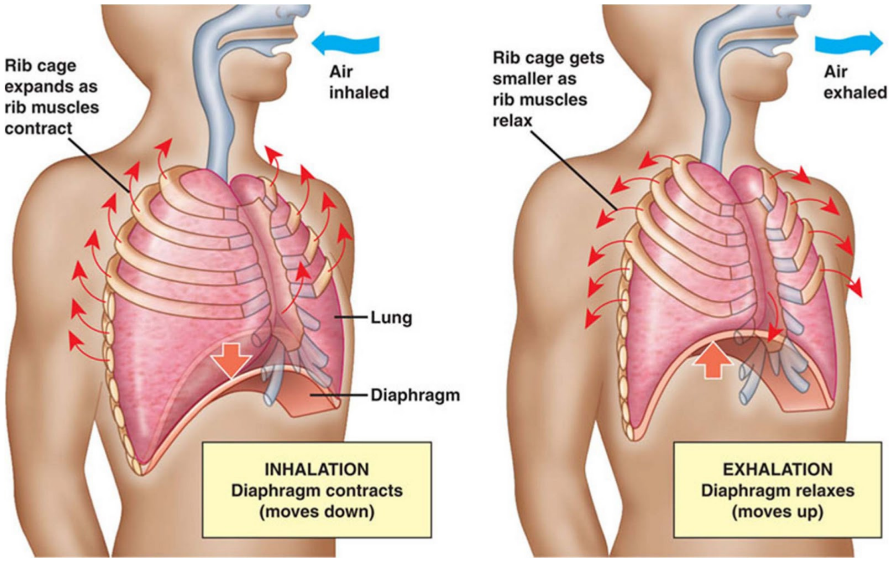


**Figure 1.1:** Diagram of the two human respiration phases: (left) inhalation and (right) exhalation.

halation is also known as a breathing (or respiratory) cycle.

Inhalation is driven by the contraction of the diaphragm, a muscle positioned in the lower part of the thorax. Its contraction causes thoracic muscles to push outwards the anterior part of the ribs and enlarge the thorax volume. The internal pressure of the thorax and inside the lungs decreases relative to the outside air pressure. The created pressure difference tends to be equalized, inducing inhalation, as the air moves from a higher pressure area to a lower one. In the exhalation phase, the lung and chest wall return to the equilibrium position and the rest shape [1]. So, the thoracic cavity volume is reduced and the air from the lungs is released. In Figure 1.1, diagrams of the inhalation and exhalation phases are depicted, showing various body parts involved in basic respiration.

Relying on the previous description of the breathing cycle, it is clear that respiration involves different parts of the human body. These movements can be analyzed in order to monitor the respiration and possibly detect specific breathing diseases.

The Respiratory Rate (RR) is an important indicator of the health of a human being; defined as the number of complete respiratory cycles carried out in a time unit, this parameter is monitored when performing clinical evaluations. It is typically measured in breath per minutes (bpm) or in breathing cycles per seconds, namely Hertz (Hz). Even if included as one of main vital signs, which include also the body temperature, blood pressure and Heart Rate (HR), the RR is often not recorded, even when the patients primary problem is a respiratory disorder [2]. Nowadays, this parameter is regularly used to monitor the progression of an illness, as an abnormal RR is an important marker of a possibly serious disease.

There is substantial evidence that alterations in the RR can be used to predict potentially serious clinical events, such as a cardiac arrest [3]. These studies have shown RR to be better than other vital measurements, such as hearth pulse and blood pressure, in discriminating between stable patients and patients at risk. Using changes in RR measurements, patients could have been identified as high risk up to 24 h before the event with a specificity of 95% [2].[1]

## 1.3 Respiratory Disorders in Newborns

### 1.3.1 Apneas

An apnea event is defined as the cessation of respiratory airflow. Traditionally, apneas in children have been considered significant if they have a duration longer than 20 s, or less if associated with desaturation, bradycardia, or arousal [4]. Apnea is more common in preterm infants. Instead, apnea is rare among full-term healthy infants and, if present, usually indicates an underlying pathology.

Thanks to newer technologies used in hospitals, the number of premature infants who survive is increasing, and therefore the number of children expe-

---

[1] The formal definition of specificity will be given in Chapter 3. Roughly, it is the fraction of negative cases correctly identified by a clinical test.

riencing apnea events is growing, too. The incidence of this life-threatening pathology is not well known. Some study accounts a rate of 2.3% of hospitalized infants, and 0.5%–0.6% of all newborns [5]. The relationship between Sudden Infant Death Syndrome (SIDS) and apnea is also of great interest: recent studies have shown that 4–13% of SIDS cases had a history of apnea occurrences [5].

Three main categories of apnea events are identified: central, obstructive, and mixed. Main features of the different categories of apneas are described in the following.

### Central Apneas

Central apneic episodes occur when there is a lack of respiratory effort due to either a cessation of output from the central respiratory centers or the inability of the efferent peripheral nerves and respiratory muscles required to receive or process the signals from the brain [6]. They may occur as result of other conditions, such as heart failure and stroke.

### Obstructive Apneas

Obstructive apneas occur when there is an obstruction in the airway and, therefore, this causes a poor or no air exchange. This type of apnea can occur when throat muscles intermittently relax and block airway during sleep. The most common sleep apneas fall in the category of obstructive apnea. The Obstructive Sleep Apnea Syndrome (OSAS) is common in childhood. Some study (in particular from Britain, Iceland, and United States) have shown similar prevalence rates of approximately 2% [4]; if untreated, OSAS may result in death. A possible link to SIDS has also been discussed [4]. A polysomnographic study of infants who died of SIDS showed an increased amount of obstructive apnea episodes, although the degree of obstruction did not appear to be clinically significant.

#### Mixed Apneas

Mixed apneas are episodes where combinations of both central and obstructive causes are involved.

### 1.3.2 Congenital Disorders

Congenital Central Hypoventilation Syndrome (CCHS) is a life-threatening rare disorder that manifests itself as sleep-associated alveolar hypoventilation, and was originally known as "Ondine's Curse." The prevalence of CCHS is unknown, but clinical registries suggest that there are between $1,000$–$2,000$ confirmed cases of CCHS worldwide; other statistics estimate a prevalence of 1 case per $200,000$ live births [7]. This syndrome, defined as the inability to automatically control the respiratory activity, is due to the absence of response to hypoxia and hypercapnia by the autonomic nervous system. Discovered as a disease caused by mutations in the paired-like homeobox 2B (PHOX2B) gene, CCHS manifests itself in the neonatal period [8]. Patients suffering from CCHS usually present events of cyanosis, respiratory failure, or apnea; it is rare that infants are subject to life-threatening events in the adulthood. Cases of SIDS caused by CCHS have also been reported. CCHS may be associated with other abnormalities, including Hirschsprung's disease, autonomic dysfunction, neural tumors, swallowing dysfunction in the childhood and minor ocular abnormalities [4]. The underlying pathophysiology of CCHS is unclear. Early detection and management of this disease with adequate ventilation and therapies help CCHS patients live fulfilling lives.

# Part I

# Monitoring of Respiration

# Chapter 2

# Monitoring of Respiration: an Overview

## 2.1 Introduction

Monitoring of physiological signals is a key tool in medical diagnostics to asses the onset and the evolution of several diseases. Among vital parameters, the Respiratory Rate (RR) holds an important role as an indicator of potential respiratory dysfunctions or changes in a patient's health condition. In particular, anomalous RR values can be a symptom of various medical conditions. For these reasons, continuous and accurate monitoring of the RR is essential. There are different approaches for respiration monitoring, but generally they can be classified as contact-based or non-contact-based. For contact-based methods, the sensing device (or part of the instrument containing it) is attached to the body of the patient. For non-contact approaches, monitoring is performed without any contact with the subject [3].

This chapter is organized as follows. In Section 2.2, methods for monitoring of respiration are introduced, first considering medical standards and contact-based methods, then recent and innovative non-contact-based systems. Within these systems, an important role is played by video processing-based methods.

In Section 2.3, an overview of the techniques for breathing monitoring and related diseases by video processing is presented. Section 2.4 is dedicated to final remarks.

## 2.2 Monitoring of Respiration and Related Diseases

The RR is a vital sign used to monitor the health condition of a patient. An abnormal RR can be an important indicator of a serious illness. There is an evidence that alterations in the RR can be used to predict potentially serious clinical events [3]. For these reasons, instruments capable to measure and monitor the RR are key tools.

In order to present an overview on various techniques to monitor respiration, it is possible to group the medical equipment into contact-based and contactless devices.

### 2.2.1 Medical and Contact-based Devices

Current gold-standard clinical monitoring methods for the RR and diseases related to respiration, in particular for sleep disorders, are based on the polysomnograph, a complex device composed of many sensors attached to the body of the patient that allows a 24 h supervision. A polysomnographic device is able to record several signals and monitor various different parameters at the same time. In order to monitor and record biological signs, a variety of sensors, electrodes and devices are employed to obtain the following measures [9].

- ElectroEncephaloGram (EEG) measures electrical activity in the brain using electrodes attached to the scalp of the patient, exploiting the fact that neurons communicate via electrical impulses. An EEG is one of the main diagnostic tests for epilepsy and is fundamental in diagnosing various brain disorders.

- ElectroCardioGram (ECG) measures the electrical activity of the heart: at each beat, an electrical impulse (or “wave”) travels through the heart; this wave causes the cardiac muscle to squeeze and pump blood. The ECG is able to record the hearth rhythm and estimate the Heart Rate (HR).
- ElectroMyoGram (EMG) is a diagnostic procedure to assess the health of muscles and nerve cells that control them (motor neurons). These neurons transmit electrical signals that cause muscles to contract.
- ElectroOculoGram (EOG) is able to record eye movements; these movements are important in determining the different sleep stages.
- Thoracic and abdominal respiratory effort and RR can be measured by elastic belts stretched around the patient’s chest and abdomen. They are able to provide an electrical waveform representing patient’s inhalation and exhalation. There are many kinds of sensors to record respiratory effort, e.g., inductive plethysmography, piezo-electric belts and mercury strain gauges.
- Pulse oximeter is an external diagnostic device that estimates the amount of hemoglobin in the blood of the patient, measuring the brightness of the light that passes through the skin. A pulse oximeter takes advantage of the red and infrared frequencies to determine the percentage of hemoglobin in the blood that is saturated with oxygen ($SpO_2$). It is typically placed on a patient’s finger.
- Capnometer is a device employed to measure the level of carbon dioxide ($CO_2$) in a sample of air. This level can provide an accurate measure of tidal flow.
- Microphones can be used for sound recording to measure snoring.

An example of polysomnographic device and its main sensors is provided in Figure 2.1.

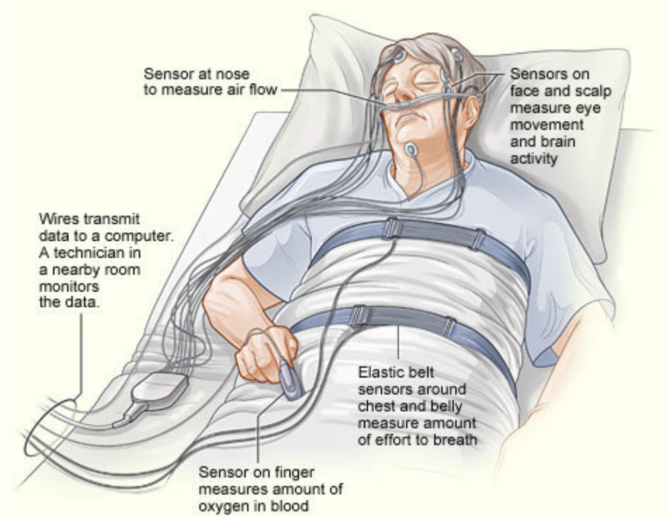


**Figure 2.1:** Diagram of a polysomnographic device and its main sensors [10].

Although these sensors and devices are able to provide accurate measurements of vital parameters, some main drawbacks may arise. First of all, these devices require many electrodes and mechanical sensors connected to the patient's body; so, the system may result slightly invasive and uncomfortable, especially for infants. The recording equipment, also because of the large number of sensors, may have a high cost and be typically available only in hospitals or clinical environments; moreover, it may not be always available for newborn patients hosted in a Neonatal Intensive Care Unit (NICU), where only ECG, $SpO_2$ saturation, blood pressure and cardiac rhythm are usually monitored. It results almost impossible to use these devices for home or remote monitoring. This technique is also time-consuming and requires long recording times for better effectiveness. In addition it cannot provide a real-time diagnosis and requires the interpretation from clinicians and specialized staff.

Recent investigations consider the development of more comfortable, wearable or portable devices to accurately measure RR, multiple vital signs and automatically detect respiratory diseases. The target is to employ modern technologies for the creation of innovative measurement devices.

### Oximetry and ECG Derived Respiration Monitoring

Oxygen saturation in the blood ($SpO_2$) is a parameter used for monitoring possible abnormal ventilation. An oximeter can simultaneously display the $SpO_2$ level as well as the pulse rate and plethysmogram. There have been studies indicating that the RR can be extracted from a plethysmogram [3,11].

Another method is to estimate the RR exploiting the fact that respiration has a modulating effect on the ECG [3,12]. By analyzing the variations of the ECG at appropriate frequencies and with suitable methods, the RR can be estimated. These methods are known as ECG-based respiration measurements. It was also demonstrated that severe respiratory dysfunctions, such as apnea events, can be detected by an analysis of the chest impedance [13], as a measure of the electrical activity in the patient's chest.

### Fiber Optic-based Devices

Optical fibers are flexible, small, transparent waveguides which are able to transmit light; these devices are used mainly for communication purposes, due to their low loss on signal propagation, their very large bandwidth and the property to be insensitive to electromagnetic interference. In particular, fiber gratings, e.g. Fiber Bragg Gratings (FBG) [14], are simple intrinsic fiber devices that can reflect, filter or disperse light passing through them. Originally used to build fiber-embedded systems for optical communications (e.g., notch filters, multiplexers, circulators), they are now used as sensor elements in different application areas.

In [15], an overview on medical applications of FBG is provided. In particular, multiple FBG in a single fiber placed in strategic locations on a bed can monitor a patient's movements. FBG underneath the patients mattress can report the pressure profile and the RR of the patient, while another set of gratings placed on top of the mattress is able to detect HR [15]. This system

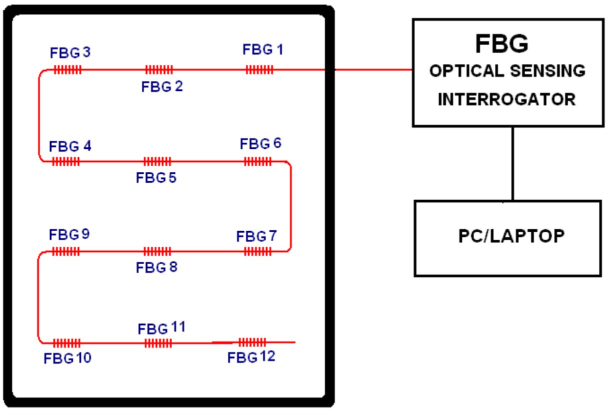


**Figure 2.2:** Schematic of a smart bed integrating optical fiber sensors for the monitoring of vital signs [15].

is a prototype also known as *smart-bed*; in Figure 2.2 an illustrative schematic of the smart-bed with FBG positions is shown.

Many devices can be built by the use of such sensors: a wearable medical textiles integrating FBG for the monitoring of respiratory movements during nuclear magnetic resonance imaging is reported in [16]. In this context, the key advantage of fiber optics is their immunity to electromagnetic fields. In [17], a device integrating FBG for the measurement of HR and RR during living activities is described: it can monitor the vibrations of human body produced by living activities—breathing and cardiac rhythm.

### Wearable Devices

The design and development of wearable sensors for health monitoring has become fundamental in scientific and industrial research. Thanks to technological progress in miniaturization of bio-sensing devices, smart textiles and optimization of low-energy wireless communication systems, the design of wearable sensor-based devices will change the future of healthcare by reducing costs and allowing focused personal health management and continuous monitoring

**Figure 2.3:** Smart jacket prototype [19].

of health conditions. So, these systems can be used as low-cost and unobtrusive solutions for 24 h health, mental and activity status monitoring.

An overview on wearable sensor-based systems for monitoring of various vital parameters is reported in [18]. As an application example, a prototype of a smart jacket for newborns able to monitor different vital signs is described in [19]. In Figure 2.3, an image of the jacket worn by a manikin for testing purposes is shown.

### 2.2.2 Non-invasive and Contactless Systems

Of significant interest is devising more comfortable, portable and low-cost vital signs and RR measurement systems, that could also allow home monitoring, reducing the time and costs of hospitalization. The biggest challenge is to find methods that do not require the application of sensors to the patients' body.

#### Microwave Sensors

Non-contact microwave sensors for sensing breathing and heartbeat have been proposed in the literature; currently research is focused on the use of two main techniques for the detection of vital signs by means of microwave sensors: continuous wave radars and ultra-wideband radars. In [20], a continuous wave

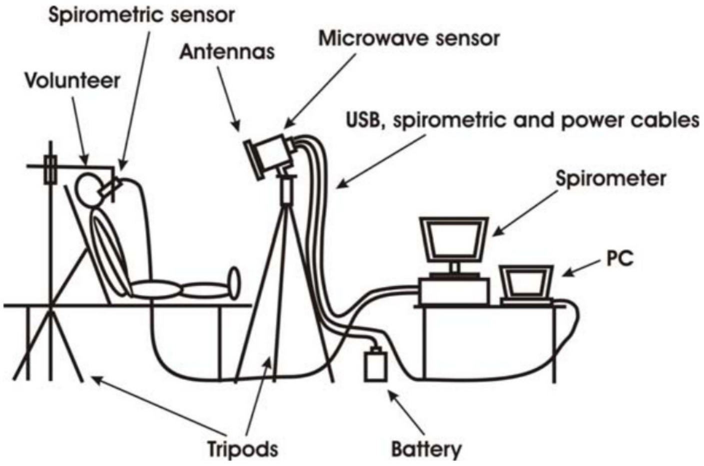


Figure 2.4: Schematic of test bench for respiration motion measurement [20].

single-frequency microwave sensor is employed to measure the back-scattered signal to detect breathing movements, with the aim to evaluate a quantitative measure of tidal lung volume on healthy patients. The sensor is a microwave coherent transceiver that radiates the patient with a monochromatic wave by directional antennas. In Figure 2.4 a schematic diagram of the measurement system by microwave sensor is depicted: data estimated by the contact-less system are compared with data of a spirometer, a device that measures the volume of inspired and expired air.

In [21] a different microwave radar for measuring vital signs in hazardous and mobile environment is proposed: this approach makes use of two kinds of radars working at different frequencies for measuring the RR and HR of a patient kept inside an isolation device. A 10 GHz radar positioned 20 cm away from the surface of the patient isolator and a 24 GHz radar placed below a stretcher underneath the patient isolator are used for the measurement of the RR and the HR, respectively.

### Ultrasonic Devices

The detection of movements and the measure of the distance between an object and a sensor can be obtained by ultrasonic devices. Accordingly, ultrasound can be employed to measure the displacement of the chest in human subjects. In [22] a 240 kHz ultrasonic proximity sensor placed 100 cm from the chest of a human subject is used to detect breathing movements and estimate the RR for long-term respiration monitoring of still patients.

### Wireless Sensors Networks

In [23] the ability to detect and monitor breathing and the RR exploiting the variations in the received signal strength measured on many links in a deployed wireless network, is proposed. Employed wireless sensors are compliant with the IEEE 802.15.4 standard, working on channel 26 with a center frequency of 2.48 GHz. The patient is supposed to be standing on line of sight links of the sensors: a sinusoidal induced variation on the received signals strength is assumed. Although a single sensor link cannot detect breathing movements, the multi-sensor data recorded by the network of devices may reliably measure the RR of the considered patient.

In Figure 2.5, a possible layout of the system based on a wireless sensor network composed of 20 sensors is shown.

## 2.3 Video Processing-based Systems

An attractive contactless diagnostic method to detect and analyze vital signs may rely on video processing-based monitoring systems, which do not require sensors connected to the patients' body. Costs related to these systems can be very low, as they depend only on the quality of the acquisition device. On the other side, complex video processing-based algorithms may be required to obtain a performance comparable to that of contact-based devices.

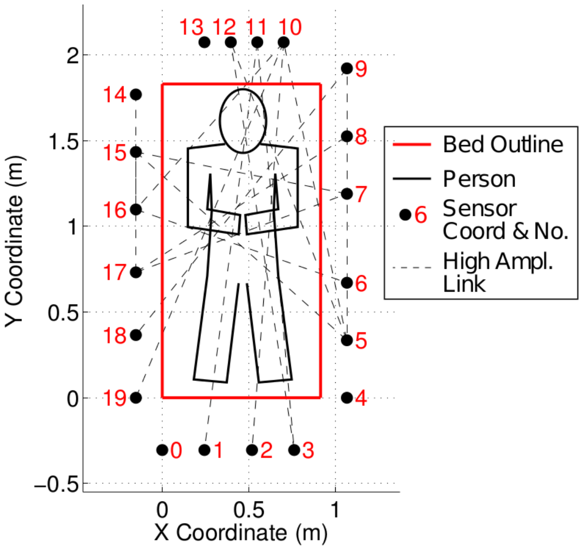


**Figure 2.5:** Layout of test bench for respiration motion measurement by a wireless sensors network [23].

### Monitoring of Pathological Movements

The development of video processing-based algorithms aimed at monitoring specific body parts for the detection of pathological movements influenced part of the work presented in this thesis about the monitoring and analysis of vital signs and respiration.

Seizures are defined as involuntary contraction of one or more muscle groups, due to a paroxysmal neuronal discharge [24]. Various types of seizures can be classified depending on the movements of body parts. In [25–27], Karayiannis *et al.* worked on video processing-based systems for detection of clonic seizures (characterized by periodic movements of the limbs), and, then, for the detection and classification of different neonatal seizures.

In more recent investigations, low-complexity algorithms for the detection of clonic seizures were proposed [28], also trying to increase detection performance by the use of multiple cameras [29,30]. Remarkable medical results were also obtained [31, 32].

### Monitoring of Vital Signs

Many systems for monitoring of a single physiological parameter or multiple vital signs based on video processing-based algorithm were proposed in the literature. They make use of various techniques or camera sensors (e.g. standard Red, Green and Blue (RGB) camera, depth camera or Microsoft Kinect [33], thermal camera, etc.) to extract information, estimate parameters, store and retrieve data. In [34], an embedded board integrated with a couple of webcams was designed to create an embedded monitor system for breath detection. The system monitors and records the respiration of the patient and sends the extracted information to a server for medical use. It can also send alarm signals in case of anomalous RR or absence of breathing movements. In [35], a method for real-time monitoring of chest and abdominal movements by a single camera was proposed. The video processing is based on simple frame differential filtering, whose drawbacks will be discussed in a following chapter. The use of a depth sensor for estimating thoracic movements and respiratory volume was also discussed in [36], where results are validated with the help of a spirometer. In Figure 2.6, the operating environment of the system described in [36] is shown. In [37], it was shown that a mobile phone can serve as a monitor for multiple physiological variables, such as HR and RR. The smart device records and analyzes the time-varying color signals of a fingertip placed in contact with its optical sensor. In [38], an algorithm for prolonged contactless monitoring of the respiration by a monochrome infrared camera was proposed and tested on healthy patients. In [39], a novel method for estimating the RR in real time from the photoplethysmogram obtained from pulse oximetry was presented. Moreover, three respiratory-induced variations (frequency, intensity, and amplitude) are extracted from the photoplethysmogram. In [40], an automatic infrared video monitoring technique for the detection of breathing anomalies and diagnosis of sleep apnea was proposed, introducing a motion model to detect subtle, cyclical breathing signals from a video obtained with a

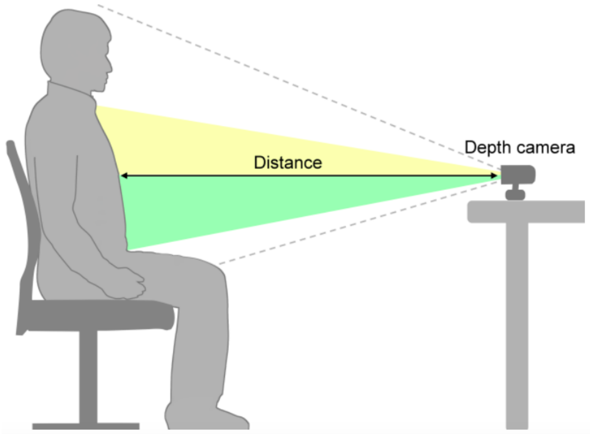


**Figure 2.6:** Setup for the measurement of respiratory volume through a depth camera [36].

low-cost 3D camera. Then, in [41] a new method for the estimation of the RR and HR by a video camera in the presence of ambient light was introduced, using autoregressive modeling. In [42], it was presented a video-based respiration monitoring method that automatically detects a respiratory region of interest and breathing signal using a single camera. The algorithm was also tested in challenging conditions and on newborn patients hospitalized in a NICU.

It is clear that recent investigations and research efforts consider the development of contactless, portable and possibly low-cost systems for the measurement of vital signs a key tool for remote non-invasive monitoring and reduction of healthcare service costs. In this context, the algorithms proposed in the following chapters of this thesis discuss various approaches for precise remote monitoring of the RR and reliable detection of apnea events.

## 2.4 Conclusion

In this chapter, an overview of the employed devices and developed methods to monitor the respiration and related disorders is provided. The approaches

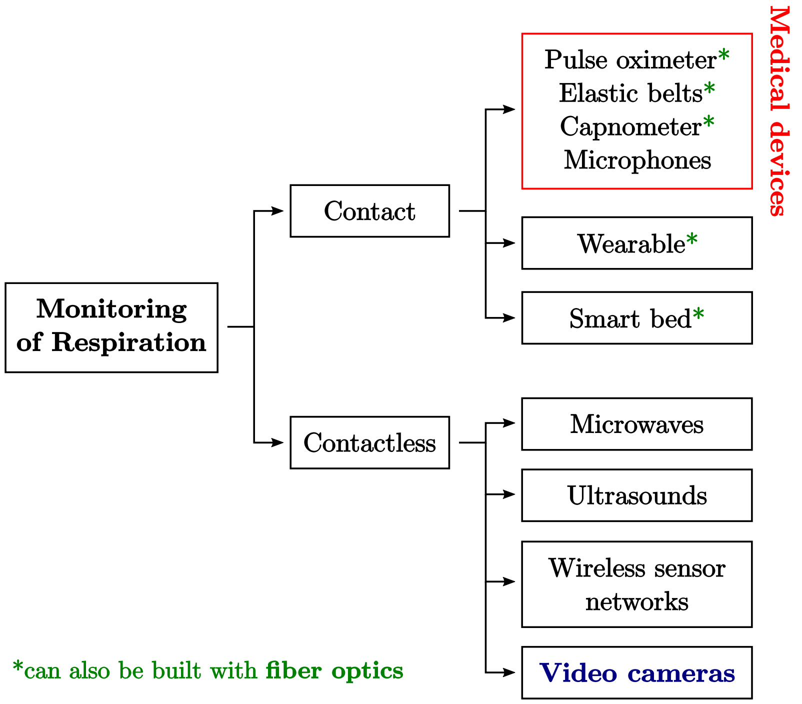


**Figure 2.7:** Diagram summarizing various methods in the literature to monitor the respiration and the RR.

are classified in two categories: contact- and non-contact-based. The devices belonging to the first category are typically used in clinical environments, can provide highly accurate measures and be built with fiber optics, which guarantee immunity to electromagnetic fields. At the same time, they have high costs and may result invasive for patients, especially for newborns. Contactless methods allow to measure vital signs in a non-invasive fashion and without sensors attached to the patient's body. In this category, appealing techniques employ camera sensors exploiting video processing-based algorithms. The methods introduced and described in this chapter are summarized in Figure 2.7. The method based on video cameras is highlighted since it is the subject of the following chapters.

# Chapter 3

# Respiration Monitoring by Video Processing

## 3.1 Introduction

In Chapter 2, the appeal and versatility of video processing-based systems as possible non-invasive measurement devices of vital signs and automatic diagnostic tools were depicted.

Monitoring of the respiratory movements, estimation of the Respiratory Rate (RR) and detection of apnea events have two main aspects in common: the motion involved during respiration has a modest extent, typically difficult to detect with classical motion detection algorithms, and is characterized by the presence (or absence, in case of apnea) of periodicity. Therefore, the ideas behind the proposed approaches for respiration monitoring consist in sensing the slight movements and applying a method to analyze the periodic behavior related to respiration and its diseases.

This chapter is organized as follows. In Section 3.2, a first method which makes use of a recent motion magnification algorithm is proposed to reveal subtle movements and detect apnea events in newborns. In Section 3.3, a second approach is developed to estimate the RR of the monitored patient by blending and optimizing the algorithms for the extraction of motion information and the motion amplification algorithm. In Section 3.4, a third method is proposed, employing a generalized model of pixel-wise periodicity and integrating an algorithm that avoids the interference of large movements not related to respiration. Finally, concluding remarks are drawn in Section 3.5.

### 3.1.1 Preliminary Definitions

A digital video can be defined as a multidimensional signal composed of sequences of digital frames, in which each frame is defined as a two-dimensional image spatially sampled and represented with a finite number of points, usually referred to as picture elements, or *pixels* [43]. Such digital images are temporally sampled with a period $T_s$ or, equivalently, with a frame rate $f_s = 1/T_s$.[1] A monochromatic (single-channel) video frame, at discrete time $n$, can be defined as a matrix $\mathbf{F}[n]$ with spatial size of $M_1 \times M_2$ pixels, where $M_1$ and $M_2$ are the number of rows and columns of the matrix, respectively. Following this notation, $F[m_1, m_2, n]$ represents the intensity of the pixel with spatial coordinates $(m_2, m_2)$ in the $n$-th video frame.

Typically, frames acquired by modern camera sensors are digital color images, in which pixels are defined in a proper color space [44]; in the following, only standard Red, Green and Blue (RGB) cameras are considered, so that each frame is composed of these three color channels. In this way, each color frame can be fully described by three matrices $\mathbf{F}_R[n]$, $\mathbf{F}_G[n]$ and $\mathbf{F}_B[n]$, which identify intensities of red, green and blue colors, respectively. Intensity values are quantized (for each channel) to a standard 8-bit representation, i.e. in the integer set $\{0, 1, \ldots, 255\}$.

A gray-scale space model is a single channel color space where only the intensity of light acquired by the camera sensor is considered. The gray scale model is extensively used for both classical image and video processing and in the following of this work, because of the simplification of the color space and

---

[1] In the following, the frame rate $f_s$ will be measured in Hz or frames per second (fps).

the reduction of data to be processed instead of the standard RGB color space. The gray scale conversion [44] from the RGB color model is here recalled and defined as

$$\mathbf{X}[n] = 0.2989\,\mathbf{F}_R[n] + 0.5870\,\mathbf{F}_G[n] + 0.1140\,\mathbf{F}_B[n] \tag{3.1}$$

where $\mathbf{X}[n]$ is the single-channel converted gray-scale frame, with 8-bit quantization.

## 3.2 Video Motion Magnification

In this section, a method to amplify subtle respiration movements is employed followed by a video processing-based system for apnea detection in newborns. The detection of respiration and apneas is based on the analysis of a signal representative of the motion framed by the video camera. At first, the extraction of motion information signal, as the one used in [28] for the analysis of larger movements, is described in Subsection 3.2.1. The motion magnification algorithm, inspired by [45] and described in Subsection 3.2.2, is introduced and then applied for monitoring of newborns.

### 3.2.1 Motion Information Extraction and Analysis

The extraction of a relevant motion signal from an RGB camera is now considered, following the approach in [28]. After gray scale conversion, a video signal $\mathbf{X}[n]$, composed of a sequence of frames with size $M_1 \times M_2$ and sampled with a time period $T_s$, is considered. In Figure 3.1, a schematic of the algorithm employed for the extraction of the motion signal is shown; each step will be described in detail in the following paragraphs.

The first step consists in a simple filtering, based on the difference between consecutive frames, also known as Difference of Frames (DoF). Accordingly, the output frame $\mathbf{D}[n]$ can be expressed as follows:

$$\mathbf{D}[n] = \begin{cases} \mathbf{0} & \text{for } n = 0 \\ \big|\mathbf{X}[n] - \mathbf{X}[n-1]\big| & \text{for } n = 1, 2, 3, \ldots \end{cases} \tag{3.2}$$

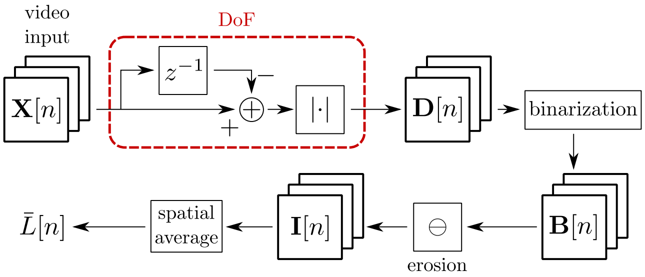


**Figure 3.1:** Block diagram of the motion signal extraction algorithm.

where $\mathbf{0}$ is a zero matrix of the same size of the original frame $\mathbf{X}[n]$. In this way, the moving parts are highlighted in the output frame sequence. In Figure 3.1, the delay of the frame $\mathbf{X}[n-1]$ is denoted with the block $z^{-1}$.

In order to further limit the complexity and reduce residual noise, the resulting frames are converted to a binary scale. This operation is described by the following simple rule

$$B[m_1, m_2, n] = \begin{cases} 1 & \text{if } D[m_1, m_2, n] > \gamma_B \\ 0 & \text{else} \end{cases} \tag{3.3}$$

where $\mathbf{B}[n]$ is the resulting binarized frame and $\gamma_B \in \{0, 1, \ldots, 255\}$ is an appropriate binary image conversion threshold. In particular, this threshold is heuristically chosen in order to maximize, in the obtained binary image, the ratio between the number of white pixels of the body parts most affected by motion and the number of remaining white pixels of the frame (including the remaining body parts), so that the obtained motion signal consists mainly of the pixels of the moving body parts [28].

After conversion to binary scale, a morphological operation called *erosion* [44] is applied, that allows the reduction of some remaining noise. After erosion, some noisy pixels may still exist but their number is negligible with respect to the number of pixels of the moving parts. Obviously, erosion leads to the loss of a portion of the signal; this loss, however, is negligible. In Figure 3.1, the erosion block is denoted with the symbol $\ominus$ [44]. The resulting frame matrix is denoted as $\mathbf{I}[n]$.

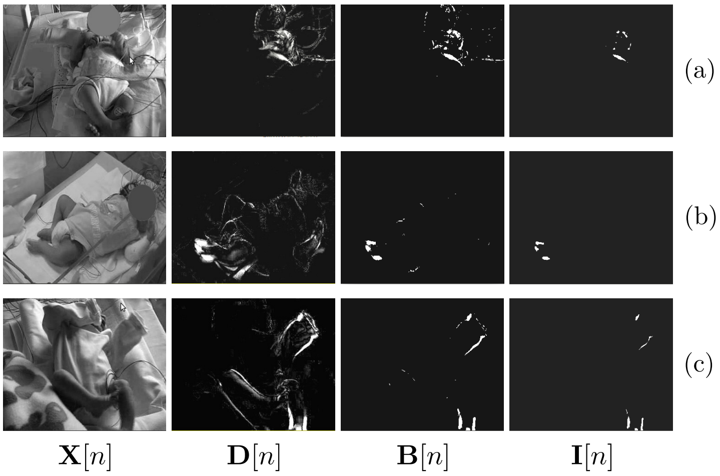


**Figure 3.2:** Motion signal extraction: gray scale, difference filtering, binary conversion and erosion for three videos, (a), (b) and (c) of newborns.

In Figure 3.2, a few illustrative examples of motion signal extraction are shown: the rows, indicated as (a), (b) and (c), are associated with different videos and the columns with the described processing steps.

Taking into account that bright binary pixels have a luminance value equal to 1, the average luminance signal, denoted as $\bar{L}[n]$, is defined as follows:

$$\bar{L}[n] \triangleq \frac{1}{M_1 M_2} \sum_{m_1=0}^{M_1-1} \sum_{m_2=0}^{M_2-1} I[m_1, m_2, n]. \tag{3.4}$$

In other words, the signal $\bar{L}[n]$ at the $n$-th frame is the fraction of white pixels in the filtered binary image. As shown in [28], this signal may represent the movement "pattern" of the involved body parts.

Once the motion signal $\bar{L}[n]$ has been extracted from an RGB camera, a method to decide on the presence of a periodic movement is needed. Since breathing fundamentally involves a quasi periodic movement, a method to determine whether the extracted signal has a periodic behavior and estimate the principal periodic component is needed. Assuming that the motion signal represents approximately the breathing motion, i.e., respiration is the only relevant body movement, the motion signal can be modeled as:

$$\bar{L}[n] = c + A\cos\left(2\pi f_0 T_s n + \phi\right) + w[n] \tag{3.5}$$

where $c$ is the continuous component, $\{w[n]\}$ is a sequence of independent and identically distributed (i.i.d.) zero-mean Gaussian noise samples, $A$ is the amplitude, $\phi$ is the phase, $f_0$ is the frequency of the periodic component, $T_s$ is the sampling period of the video sequence and $n$ is the frame index. The main goal is to detect the presence of the sinusoidal component and estimate $f_0$.

In order to estimate $f_0$, the Maximum Likelihood (ML) criterion can be applied. In [46], given the vector of parameters $\boldsymbol{\theta} = [A, f_0, \phi]$, the ML estimation strategy of $\boldsymbol{\theta}$ is shown to consist of the minimization of the following likelihood function:

$$J\left(\boldsymbol{\theta}\right) = \sum_{n=0}^{N-1} \left(\bar{L}[n] - A\cos\left(2\pi f_0 T_s n + \phi\right)\right)^2 \tag{3.6}$$

where $N$ is the number of analyzed samples. After standard algebraic manipulations and approximations [46], the ML frequency estimate becomes

$$\hat{f}_0 = \underset{f\in\mathcal{F}}{\operatorname{argmax}} \left| \sum_{n=0}^{N-1} \bar{L}[n] e^{-j2\pi f T_s n} \right|^2 = \underset{f}{\operatorname{argmax}} \left|\mathrm{DFT}\left\{\bar{L}\left[n\right]\right\}\right|^2 \tag{3.7}$$

where DFT$\{\cdot\}$ and $\mathcal{F}$ denote the Discrete Fourier Transform (DFT) operator and the set of search frequencies related to respiration, respectively. It is possible to interpret the argument of the argmax$(\cdot)$ operator in (3.7) as the expression of the periodogram of $\bar{L}$ and, thus, its peak indicates the most relevant harmonic and corresponds to the estimated frequency.

Once the frequency has been estimated, the presence of a significant periodic component has to be determined. To this end, the estimate $\hat{A}$ of the amplitude component can be computed as follows:

$$\hat{A} = \frac{2}{N} \left| \sum_{n=0}^{N-1} \bar{L}[n] e^{-j2\pi \hat{f}_0 T_s n} \right|. \tag{3.8}$$

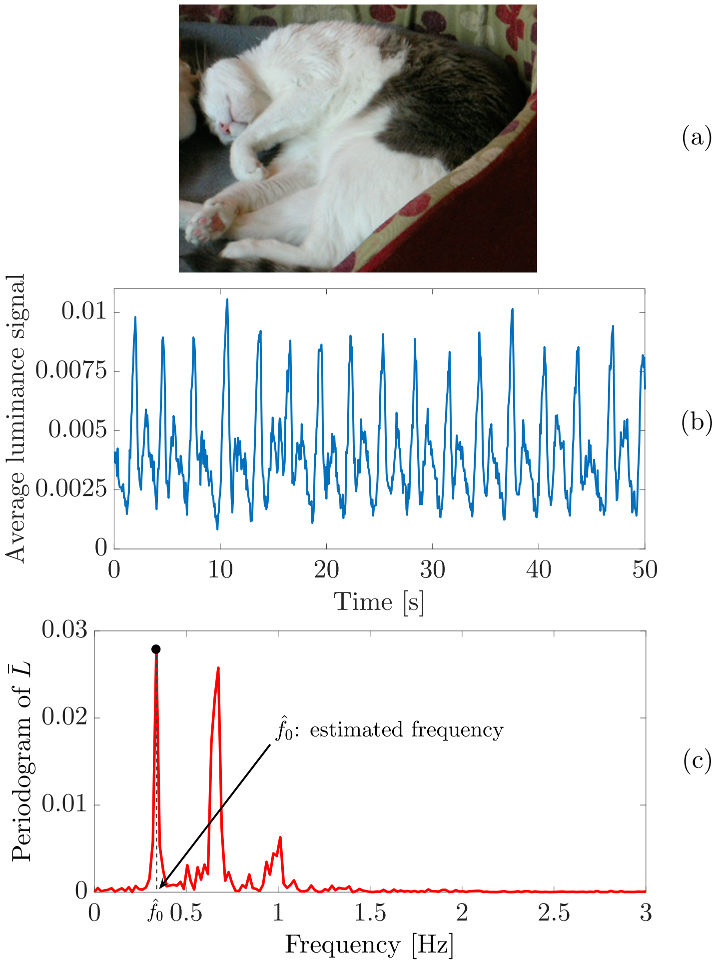


**Figure 3.3:** (a) Image acquired from a video camera framing a sleeping cat, (b) example of a breathing signal extracted by video processing and (c) corresponding periodogram from the likelihood function with frequency estimation.

A significant periodic component is detected if [23]:

$$N\hat{A}^2 > \eta \tag{3.9}$$

where $\eta$ is an appropriate threshold experimentally determined.

In Figure 3.3(a) an image of a video recording, sampled at 15 Hz, framing a sleeping cat is shown. In Figure 3.3(b) an example of average luminance signal

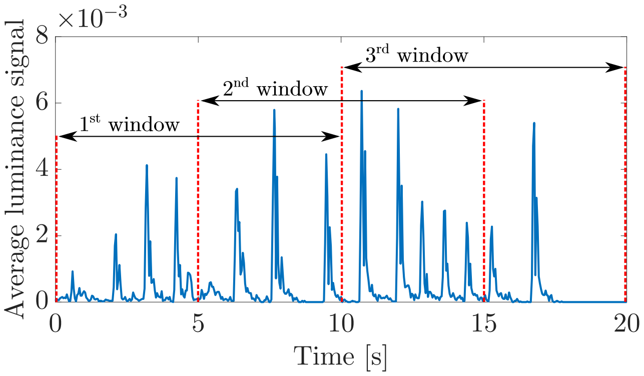


**Figure 3.4:** Examples of interlaced observation windows of video frames for periodicity analysis.

$\bar{L}[n]$ extracted from the video of the cat is exhibited; a periodic component in the signal is clearly visible.[2] In Figure 3.3(c) the corresponding periodogram and the estimated frequency $\hat{f}_0$ in correspondence of the peak are shown.

The described ML method can be applied on length-$N$ (in terms of frames) disjoint observation windows; in other words, the periodicity can be analyzed window by window. However, in order to improve the reliability of the proposed method, the extracted average motion signal can be analyzed on half-interlaced consecutive windows. In fact, an apnea could manifest itself across two consecutive disjoint windows: a single window-based algorithm could thus miss it in each of the two windows. In Figure 3.4, an illustrative example of interlaced window observations relating to quasi periodic movements in the context of neonatal seizure detection is shown [28].

### 3.2.2 Eulerian Video Magnification

Detection of subtle motion for clinical analysis is a remarkable problem in biomedical monitoring based on video cameras. In fact, small movements can be confused with noise or small environmental changes in the scene. A solu-

---

[2] The average luminance signal is extracted using the motion magnification algorithm described in the following Subsection 3.2.2.

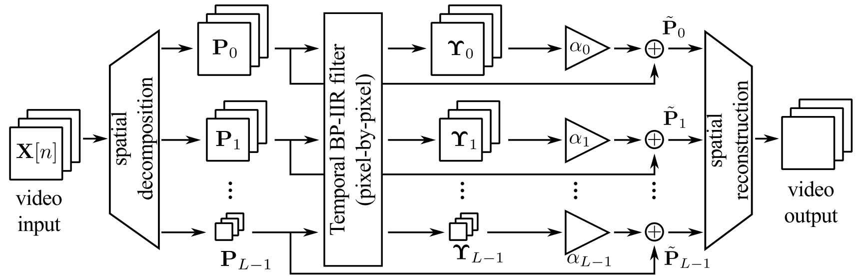


**Figure 3.5:** Block diagram of the EVM algorithm for motion amplification.

tion for detecting these movements was proposed in [45], where the presented algorithm can overcome spatio-temporal limitations of the human visual system. In the following, the video signals are processed by the approach in [45], named as Eulerian Video Magnification (EVM), in order to amplify respiratory motion in the proposed monitoring system.

The EVM algorithm enhances small changes in the video by a cascade combination of a frame-by-frame spatial processing and a temporal processing on the pixel intensity variation [45]. In Figure 3.5, a block diagram of the EVM algorithm is presented. The EVM approach consists of four main steps, described in the following subsections.

#### 3.2.2.1 Spatial Multi-scale Decomposition

The first step decomposes every video frame into a set of images, each representing different visual details. These details correspond to different spatial frequency bands of the frames. Decomposition is performed through computation of an image pyramid, namely a hierarchical representation of the original image $\mathbf{X}[n]$ by a set of $L$ images, called levels, each with different resolution. For this reason, an image pyramid is also called multi-scale decomposition. The employed decomposition, sketched in the first block of Figure 3.5, is the Laplacian pyramid [47], which splits each frame into scaled images representing different spatial frequency sub-bands. The levels of the Laplacian pyramid are sorted from a lower level (with higher resolution, more details, and representing high spatial frequency bands) to a higher level (with lower resolution, coarser details and representing low spatial frequency bands). Each scaled image that composes the levels of the Laplacian pyramid is obtained starting from a combination of levels of a different pyramidal decomposition: the Gaussian pyramid [47].

An $L$-level Gaussian pyramid decomposition is obtained starting from the original image $\mathbf{G}_0[n] = \mathbf{X}[n]$, set as the lower level of the pyramid, and deriving the upper levels, specified by matrices $\mathbf{G}_1[n], \mathbf{G}_2[n], \ldots, \mathbf{G}_{L-1}[n]$, with a sequential process of spatial low-pass filtering and downsampling. The levels of the pyramid can be recursively computed by a "reduce" function defined as follows

$$\begin{aligned}\mathbf{G}_\ell[n] &= \mathrm{RED}\left(\mathbf{G}_{\ell-1}[n]\right) \\ &= \sum_{k_1=-R_M}^{+R_M} \sum_{k_2=-R_M}^{+R_M} h_M[k_1, k_2] G_{\ell-1}\left[2m_1 - k_1, 2m_2 - k_2, n\right]\end{aligned} \tag{3.10}$$

in which $\ell = 1, 2, \ldots, L-1$, $\mathrm{RED}(\cdot)$ is the reducing function, $h_M[m_1, m_2]$ is a proper truncated Gaussian low-pass filter mask [47] of size $2R_M + 1 \times 2R_M + 1$ (size of the two-dimensional filter), $R_M$ is a positive integer and $\mathbf{G}_{\ell-1}[n]$ is the previous level of the pyramid. A reverse operator $\mathrm{EXP}(\cdot)$ can also be introduced by interpolating and "expanding" the $\ell$-th level to the same size of the lower one, defining

$$\begin{aligned}\hat{\mathbf{G}}_\ell[n] &= \mathrm{EXP}\left(\mathbf{G}_{\ell+1}[n]\right) \\ &= 4 \sum_{k_1=-R_M}^{+R_M} \sum_{k_2=-R_M}^{+R_M} h_M[k_1, k_2] G_{l+1}\left[\frac{m_1 - k_1}{2}, \frac{m_2 - k_2}{2}, n\right]\end{aligned} \tag{3.11}$$

where the filter mask $h_M[m_1, m_2]$ is the same as in (3.10).

Laplacian pyramid levels are defined as error images [47] between a level of the Gaussian pyramid and the lower one properly expanded by the function defined in (3.11). More formally, the $\ell$-th level of the Laplacian pyramid is an

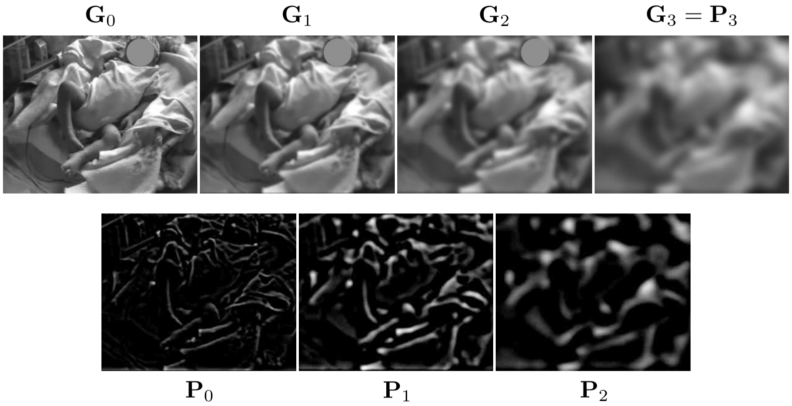


**Figure 3.6:** Illustrative examples of expanded Gaussian pyramid (first row) and expanded Laplacian pyramid (second row) for $L = 4$.

image defined as

$$\begin{aligned}\mathbf{P}_\ell[n] &= \mathbf{G}_\ell[n] - \mathrm{EXP}\left(\mathbf{G}_{\ell+1}[n]\right)\\ &= \mathbf{G}_\ell[n] - \hat{\mathbf{G}}_\ell[n]\end{aligned} \tag{3.12}$$

where $\ell = 0, 1, \ldots, L-2$. The last level of the pyramid is defined as $\mathbf{P}_{L-1}[n] = \mathbf{G}_{L-1}[n]$. Illustrative examples of Gaussian and Laplacian pyramids are shown in Figure 3.6. Each level image is a different band-pass spatially filtered image and shows different details of the original image.

#### 3.2.2.2 Temporal Filtering

In this step, a pixel-wise processing is performed to extract pixel intensity variations of interest. In the context of the apnea detection system, this step is used to select periodic movements in the video which are useful for breathing detection. Band-pass digital filters are applied along the time dimension of the video signal. In order to maintain the computational complexity low, still using sufficiently selective temporal filters, a band-pass filter obtained by the difference of two low-pass Butterworth Infinite Impulse Response (IIR) digital second-order filters with different cutoff frequencies $f_{c,L} < f_{c,H}$ is employed. Defining the difference equation of the IIR low-pass filter with a generic cutoff frequency $f_c$ on a single level $\ell$ of the pyramid at discrete time $n$ as

$$\begin{aligned}\boldsymbol{\Upsilon}_{\ell,f_c}[n] =& \frac{1}{a_0}\Big[b_0\mathbf{P}_\ell[n] + b_1\mathbf{P}_\ell[n-1] + b_2\mathbf{P}_\ell[n-2] \\ & - a_1\boldsymbol{\Upsilon}_{\ell,f_c}[n-1] - a_2\boldsymbol{\Upsilon}_{\ell,f_c}[n-2]\Big]\end{aligned} \tag{3.13}$$

where the coefficients $a_0$, $a_1$, $a_2$, $b_0$, $b_1$ and $b_2$ of the digital filter are properly selected taking into account the cutoff frequency $f_c$ and the frame rate $f_s$ of the video. The overall filtered band-pass frames can be expressed as

$$\boldsymbol{\Upsilon}_\ell[n] = \boldsymbol{\Upsilon}_{\ell,f_{c,H}}[n] - \boldsymbol{\Upsilon}_{\ell,f_{c,L}}[n]. \tag{3.14}$$

The temporal filtering method outlined above is computed on every level of the Laplacian pyramid with equal tap weights in the digital IIR filters.

#### 3.2.2.3 Variable Gain Amplification

After temporal filtering on spatially decomposed frames, the EVM algorithm amplifies selected movements. Amplification is obtained by multiplying the filtered pyramid levels $\{\boldsymbol{\Upsilon}_\ell\}_{\ell=0}^{L-1}$ by a proper amplification factor, in order to enhance moving parts of interest. Each $\ell$-th pyramid level must be amplified with a different coefficient $\alpha_\ell$, set to avoid visual artifacts [45]. Considering that movements can be defined as spatial displacements of framed objects (or pixels) varying during time, combined spatial and temporal processing leads to movement enhancement thanks to a first-order Taylor series expansion of the image intensity signal and the use of a proper definition of the displacement function. Every amplification factor is function of the radial spatial wavelength: however, as a result of a first-order approximation of motion and under the assumption of smooth images and small motions in the video, the amplification coefficients must fulfill a bound whereby they are maintained constant for higher spatial wavelengths and are linearly attenuated for lower ones [45]. So, in order to avoid noise amplification or motion artifacts, each level of the multi-scale decomposition is amplified with different coefficients $\{\alpha_\ell\}_{\ell=0}^{L-1}$, which take properly set values at higher levels and decreasing ones at lower levels.

#### 3.2.2.4 Reconstruction of Video Frames

Amplified levels are now summed to the original levels of the pyramid (to enhance motions) and are combined to reconstruct a single video output. This video signal output with amplified respiration movements can then be used, through ML processing with interlaced windows shown in Subsection 3.2.1, to detect breathing (or, more appropriately, absence of breathing, i.e., apnea). Following [45], the algorithm outputs frames which reconstruct the processed Laplacian pyramid. Output frames are calculated with an iterative system starting from the $L$-th level of the pyramid and rebuilding the original level. This iterative reconstruction can be expressed as follows:

$$\tilde{\mathbf{P}}_\ell[n] = (\alpha_\ell \boldsymbol{\Upsilon}_\ell[n] + \mathbf{P}_\ell[n]) + \mathrm{EXP}\left(\tilde{\mathbf{P}}_{\ell+1}[n]\right) \tag{3.15}$$

starting from $\ell = L - 1$ and ending with $\ell = 0$.

In Figure 3.7, the luminance signal obtained without and with breath magnification is shown. The two signals refer to a 18 s video, recorded at $f_s = 25$ Hz, in which an infant, although breathing, seems still at naked eye. In Figure 3.7(a), in the extracted motion signal $\bar{L}[n]$ (without EVM magnification) small breathing movements are indistinguishable from background noise. In Figure 3.7(b), the motion signal $\bar{L}[n]$, based on the use of EVM in order to emphasize breathing, is shown: the periodicity of the small breathing movements is here visible. A frequency of about 0.5 breath/s can be estimated.

In Figure 3.8, an illustrative example of the image amplification, for breathing detection, obtained by the use of EVM is shown.

### 3.2.3 Breathing Analysis and Apnea Detection

Recall that an apnea event is defined as an episode of absence of breathing lasting at least 20 s, or between 10 and 20 s provided that it is associated with other clinical signs/symptoms. Once a respiratory movement has been emphasized, the extracted motion signal can be analyzed with the aim of identifying the time intervals in which breathing is not detected. In Figure 3.9,

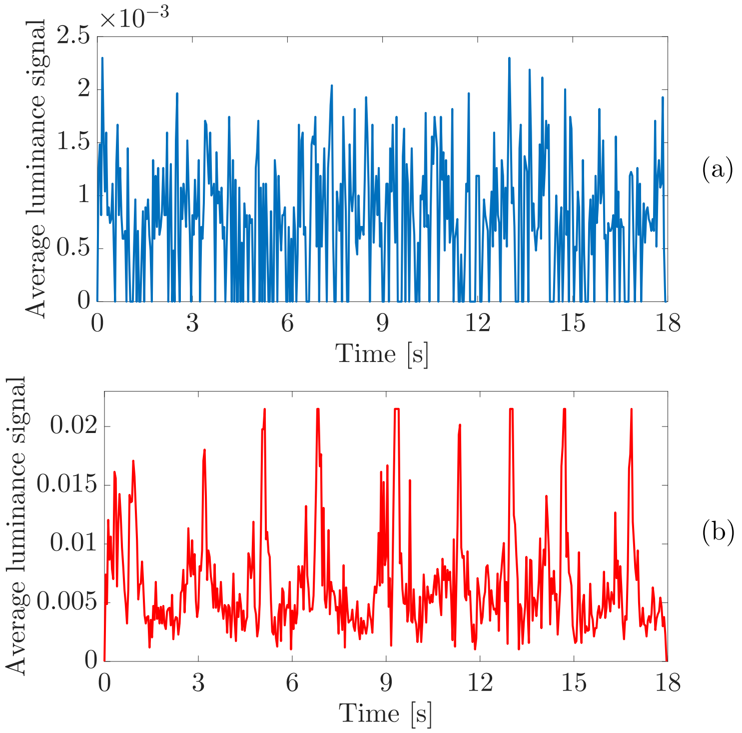


**Figure 3.7:** Examples of extracted breathing signals: (a) without pre-magnification and (b) with pre-magnification.

(a) (b)

**Figure 3.8:** Emphasizing the respiratory movements: (a) original image (b) image with enhanced breathing motion.

the motion signals extracted from two video sequences, each lasting 30 s, with or without emphasizing breathing, are compared. In Figure 3.9(b), a threshold is employed in order to discriminate the residual noise; a value of $2.5 \cdot 10^{-3}$ was heuristically selected for this threshold. As it can be observed, the proposed method highlights the time frame in which breathing movements are absent.

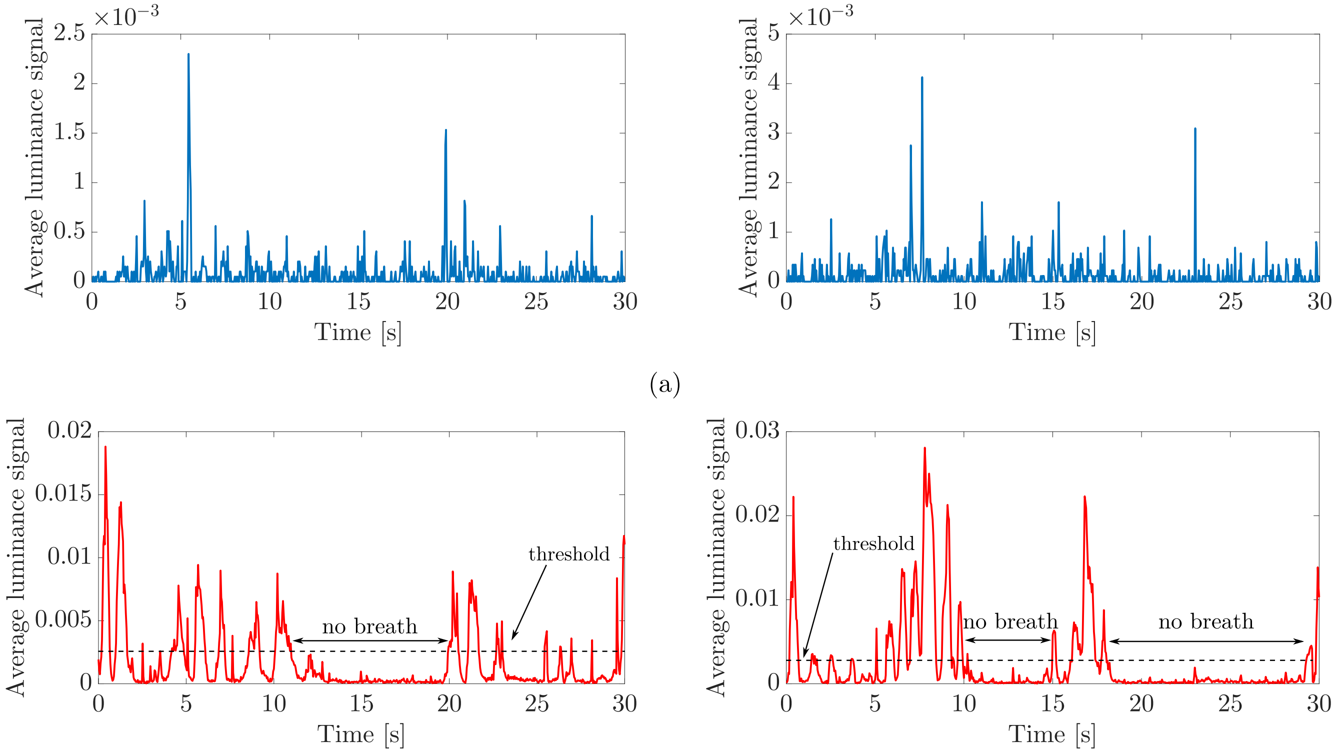
### 3.2.4 Performance Analysis

The performance for detection of apneas of the previously described algorithm is now discussed. The system is tested on videos recorded with frame rate $f_s = 25$ Hz and set up with half-interlaced windows of duration $NT_s = 20$ s for the analysis of breathing movements. Regarding the EVM algorithm, it is employed with $L = 6$ levels, cutoff frequencies of the temporal filters set to $f_{c,L} = 0.25$ Hz and $f_{c,H} = 1.1$ Hz and maximum amplification factors in the pyramidal levels belonging to the range $[10, 30]$. Afterwards, the extraction of the average luminance signal is performed, in the method shown in Figure 3.1, using a binarization threshold $\gamma_B$ equal to 10. Results of a retrospective analysis performed on the video of a newborn who previously received a timely diagnosis of central apneas by means of a polysomnography, are now presented. As ground truth, the results obtained by the video-ElectroEncephaloGram (EEG) polygraphy with thermistor and pletismography, lasting a total of 1 h and 35 s are considered. They indicate[3] 23 episodes with a total duration of 28 min and 31 s, an average single apnea duration of 74 s, a maximum duration of 3 min and 25 s and a minimum duration of 11 s.

In analyzing the performance of the proposed automatic diagnostic system, only 17 of the 23 apnea events detected by the polysomnograhic EEG system are considered, because in 4 apnea episodes, the view of the newborn was obstructed by medical staff and in another 2 episodes, the apnea events were too short and without other clinical signs, whereas the system has been

---

[3] Due to the documented dysregulation of respiratory control, during the recording the patient was also monitored by an elastic belt sensor, a nasal flow-meter and a pulse oximeter. Furthermore, oxygen therapy was applied using nasal cannulas whenever clinically indicated.



(a)

(b)

**Figure 3.9:** Two examples of breathing signals: (a) without pre-magnification and (b) with pre-magnification and thresholding.

calibrated to detect longer respiratory pauses.

The performance of the proposed system is investigated in terms of sensitivity and specificity, defined, respectively, as follows:

$$\begin{aligned} \alpha &= \frac{T_{\mathrm{TP}}}{T_{\mathrm{TP}} + T_{\mathrm{FN}}} \\ \beta &= \frac{T_{\mathrm{TN}}}{T_{\mathrm{TN}} + T_{\mathrm{FP}}} \end{aligned} \tag{3.16}$$

where $T_{\mathrm{TP}}$, $T_{\mathrm{TN}}$, $T_{\mathrm{FP}}$ and $T_{\mathrm{FN}}$ denote, respectively; the total length of the time intervals with apnea correctly detected (time of True Positives); the total length of the time intervals with no apnea correctly detected (time of True Negatives); the total length of the time intervals with regular breathing incorrectly reported as apnea (time of False Positives) and the total length of the time intervals with apnea incorrectly reported as normal breathing (time of False Negatives). A sensitivity (in percentage) of 100% means that the test recognizes all the time intervals $T_{\mathrm{TP}}$ (i.e, all apneas in our case) and no $T_{\mathrm{FN}}$. Thus, a negative result in a high sensitivity test excludes the disease. A specificity (in percentage) of 100% means that the test recognizes all the time intervals $T_{\mathrm{TN}}$ (i.e., all breathing movements) and no $T_{\mathrm{FP}}$. Since 100% specificity means that no positives are erroneously tagged, a positive result in a high specificity test strongly indicates the presence of the pathology. In other words, while the sensitivity quantifies the test ability to correctly detect the disease, the specificity is a measure of the test ability to correctly exclude the disease.

In the evaluation of these time intervals, a Tolerance Delay (TD), defined with respect to the end of the first processing window where the apnea episode can be detected, is considered [30]. An apnea episode is assumed correctly detected if the algorithm reports it within a time interval not larger than the assigned TD. Figure 3.10 helps to better understand this concept, where a long apnea is considered and consecutive interlaced windows are shown: various TDs are shown jointly with consecutive observation windows. As TD increases, a larger number of interlaced windows fall inside the TD, thus increasing the

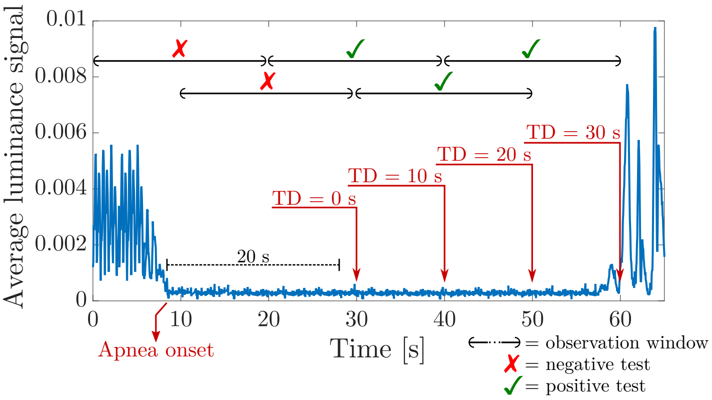


**Figure 3.10:** Half-interlaced observation windows of 20 s length, with TD of 10, 20 and 30 s, apnea starting at 8 s and lasting approximately 47 s.

probability of correctly detecting the presence of an apnea. In the upper part of Figure 3.10, examples of detection results are shown; for TD equal to 0 s, the apnea episode is missed, whereas for TD equal to 10, 20 or 30 s, it is correctly detected.

In Table 3.1, the sensitivity and specificity values obtained with the proposed algorithm are shown, considering various values of the assumed TD. In particular, $\alpha$ and $\beta$ are computed by considering the portions of minutes of apnea (out of 22 min) and regular breathing (out of 38 min) correctly/incorrectly diagnosed. Since apnea episodes are potentially deadly, it is important to promptly detect them. The goal is to identify apnea events immediately upon cessation of breathing. The proposed system was able to detect 13 to 17 episodes, out of a total of 17, depending on the acceptable delay in reporting an episode. The analysis was performed on temporal windows of 20 s.

The analyzed video was recorded in the Neonatal Intensive Care Unit (NICU) of the University Hospital of Parma. In this context, the movement of medical personnel may cause continuous changes in brightness, because, although not framed by the camera, they may be reflected on the plexiglas

| **Performance in Apnea Detection** | | | | | | | |
|---|---|---|---|---|---|---|---|
| TD | DA | $T_{\mathrm{TP}}$ | $T_{\mathrm{TN}}$ | $T_{\mathrm{FP}}$ | $T_{\mathrm{FN}}$ | Sensitivity | Specificity |
| 0 | 13 | 1200 | 1800 | 500 | 140 | 90% | 78% |
| 10 | 14 | 1230 | 1800 | 500 | 110 | 91% | 78% |
| 20 | 16 | 1280 | 1880 | 420 | 60 | 95% | 81% |
| 30 | 17 | 1340 | 1920 | 380 | 0 | 100% | 83% |

Legend: TD = Tolerance Delay (seconds); DA = number of Detected Apneas; $T_{\mathrm{TP}}$, $T_{\mathrm{TN}}$, $T_{\mathrm{FP}}$ and $T_{\mathrm{FN}}$ (seconds).

**Table 3.1:** Sensitivity and specificity performance on real apnea detection in newborns.

cradles. As the proposed algorithm is sensitive to variations in brightness, an improved lighting system would certainly lead to a performance improvement.

## 3.3 Spatio-Temporal Motion Analysis

This section describes an algorithm for RR estimation based on the extraction of breathing motion information using multi-scale decomposition, proper temporal filtering and motion signal extraction, followed by ML estimation. Experimental results, obtained by comparing the ML estimate by the proposed system with the gold-standard polysomnographic system, showing a good agreement, are presented.

Recall that breathing is fundamentally characterized by a periodic movement of specific body parts, among which the chest is the most evident. In order to perform ML estimation, the monitoring system requires the extraction of a breathing motion signal by means of video-based processing. Once a signal representative of breathing is obtained, the ML estimate is approximated using a periodicity estimator.

To handle the very small breathing movements, especially of a newborn, the following proposed method takes advantage of a spatio-temporal processing,

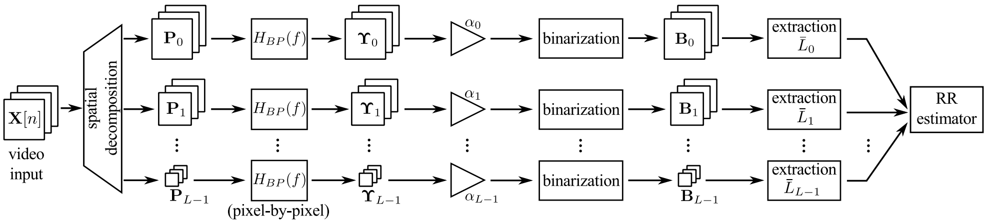


**Figure 3.11:** Block diagram of the spatio-temporal RR estimation technique.

similar to the one presented in Section 3.2, extracting multiple motion signals which are used for robust RR estimation. Unlike the previous EVM-based method, the proposed algorithm does not perform the final reconstruction of the video stream and employs a slightly different method for temporal filtering and motion signal extraction. In Figure 3.11, a block diagram of the proposed system is shown: the various blocks are described in detail in the following subsections.

### 3.3.1 Breathing Information Extraction

The first step has the goal to extract breathing motion signals from a video of patient at rest. The algorithm integrates the spatial and the temporal processing included in the EVM with direct motion signal extraction on each level of the spatial decomposition. The whole processing is performed on a gray scale video $\mathbf{X}[n]$ sampled with rate $f_s$, obtained as described in Subsection 3.1.1.

#### 3.3.1.1 Multi-scale Decomposition

The first processing step decomposes every video frame into a set of images, each representing different visual details. These details correspond to different spatial frequency bands of the frames. Decomposition is performed through computation of an image pyramid. The employed decomposition, sketched in the first block of Figure 3.11, is the Laplacian pyramid [47] employed in the EVM algorithm and introduced in detail in Subsection 3.2.2.

#### 3.3.1.2 Temporal Filtering and Motion Amplification

After the multi-scale decomposition, a pixel-wise processing is applied to every level of the pyramid to extract pixel intensity variations related to breathing motion. Due to the strong periodic behavior of breathing movements, this step uses a digital filter applied to the temporal dimension of the video signals to select periodic variations in the scene in a proper frequency band. To keep the computational complexity low and reduce the delay introduced by filtering, still maintaining sufficient selectivity, a second-order digital band-pass IIR filter is used. This digital band-pass filter, with frequency response $H_{BP}(f)$, is applied to all pyramid levels, as shown in Figure 3.11. Accordingly, the system processes video frames by a Butterworth IIR filter, properly designed by setting up cutoff frequencies $f_{c,L}$ and $f_{c,H}$, with $f_{c,L} < f_{c,H}$. Butterworth filters are characterized by the property that the magnitude response is maximally flat in the pass-band [48] and 3 dB below its maximum value at the cutoff frequencies $f_{c,L}$ and $f_{c,H}$. The cutoff frequencies of the temporal filter applied to all levels of the pyramid are properly selected to extract periodic variations related to breathing movements. Thus, given the order of the filter, the cutoff frequencies and the sampling frequency $f_s$, the coefficients of the digital filter are readily obtained [48, 49].

As in the EVM algorithm, once every level of the Laplacian pyramid has been filtered, it is possible to amplify selected levels. This procedure is obtained multiplying each filtered pyramidal level $\boldsymbol{\Upsilon}_\ell$, shown as filter output frames in Figure 3.11, by a proper amplification factor to enhance only periodic movements of interest. In order to avoid noise amplification or motion artifacts, each level of the multi-scale decomposition is amplified with different coefficients $\{\alpha_\ell\}_{\ell=0}^{L-1}$. Indeed, it could be shown that amplification coefficients are strictly related with the spatial frequencies, due to the Eulerian motion model and the spatio-temporal relationship shown in [45]. The coefficients $\{\alpha_\ell\}_{\ell=0}^{L-1}$ are set as follows: much larger than 1 for higher pyramidal levels (lower spatial frequency bands) and linearly attenuated for lower levels (higher spatial frequency bands) down to $\alpha_0 = 1$. The value of the coefficient employed for higher

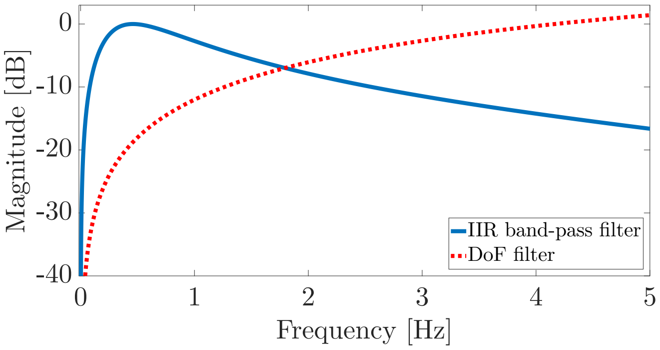


**Figure 3.12:** Magnitude frequency response of the DoF FIR filter and the second-order band-pass IIR filter used for breathing motion extraction.

pyramidal levels is heuristically chosen, depending on how much amplification of movements is needed in order to properly detect breathing movements. The last filtered pyramidal level $\boldsymbol{\Upsilon}_{L-1}$ is not used because of too low resolution and, correspondingly, gain $\alpha_{L-1}$ is set to zero.

#### 3.3.1.3 Extraction of Motion Signals

Before estimating the RR, the system needs to extract a signal that represents the quantity of motion present in the scene. Motion signal extraction is performed on every level of the filtered and amplified Laplacian pyramid. In the algorithm in Section 3.2, the EVM algorithm was used as a video preprocessing tool to magnify small motions: then, relying on the motion signal extraction algorithm described in Subsection 3.2.1, the capability to detect respiratory pauses, or apnea episodes, was demonstrated. However, the extracted signal are also corrupted by noise and it is quite difficult to obtain a precise estimation of the RR. Motion signal extraction in [28] detects scene variations by the use of a well known algorithm based on a simple differential filtering method, called DoF, which is described by the equation (3.2). In Figure 3.12, the frequency response of the DoF digital filter is shown, considering a sampling frequency of $f_s = 25$ Hz; it is clear that the DoF method corresponds to a high-pass filter able to extract fast variations. As clearly visible, the DoF filter is not appropriate to extract breathing motions, particularly if they are not amplified. In fact, newborn respiration acts have typical periodicity in the frequency range between 0.25 Hz and 1 Hz. Since the magnitude frequency response of the DoF filter in the range between 0.25 Hz and 1 Hz is always below $-10$ dB, variations of interest are deeply attenuated.

To solve this problem and combine the motion signal extraction with the spatio-temporal processing, the proposed system uses of the band-pass filter previously introduced and integrated in the EVM algorithm instead of the DoF filter to extract motions. The pixel-wise digital filter has transfer function of type:

$$H_{BP}(z) = K_{BP} \frac{\left(1 + z^{-1}\right)\left(1 - z^{-1}\right)}{\left(1 - pz^{-1}\right)\left(1 - p^{*}z^{-1}\right)} \tag{3.17}$$

where $K_{BP}$ is a scale factor, $p$ and $p^*$ are the complex conjugate poles, properly determined by the filter design rules to fit requirements for $f_{c,L}$ and $f_{c,H}$ [48, 49]. The filter used for breathing analysis is set up with $f_{c,L} = 0.2$ Hz and $f_{c,H} = 1.05$ Hz, so that normal periodic breathing movements are included. In Figure 3.12, the frequency response of the used IIR digital filter is also shown.

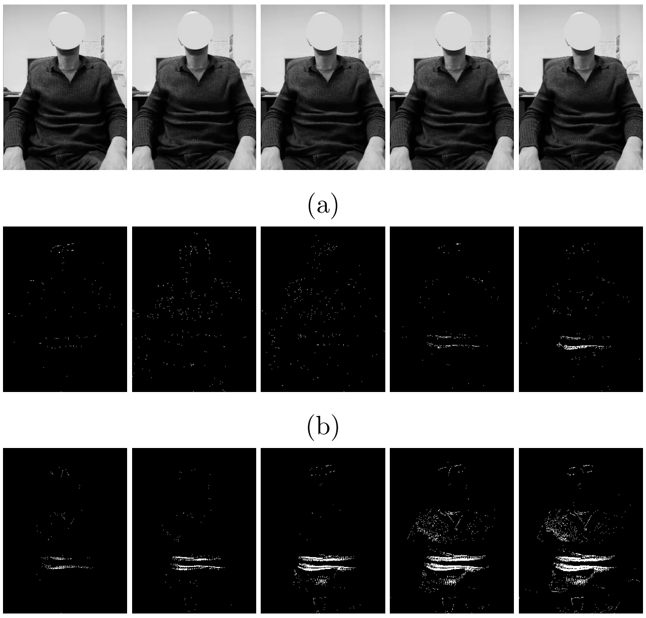
Filtered and amplified outputs $\{\alpha_\ell \boldsymbol{\Upsilon}_\ell\}_{\ell=0}^{L-1}$ are now processed for motion signal extraction. In order to limit the computational complexity and reduce noise, the resulting amplified frames are converted to binary scale. This operation, performed in the binarization blocks of Figure 3.11, can be described as follows:

$$B_\ell[m_1, m_2, n] = \begin{cases} 1 & \text{if } |\alpha_\ell \Upsilon_\ell[m_1, m_2, n]| \geq \gamma_B \\ 0 & \text{otherwise} \end{cases} \tag{3.18}$$

where $B_\ell[m_1, m_2, n]$ and $\Upsilon_\ell[m_1, m_2, n]$ denote the entries of matrices $\mathbf{B}_\ell[n]$ and $\boldsymbol{\Upsilon}_\ell[n]$ at position $(m_1, m_2)$, $\mathbf{B}_\ell[n]$ is the binarized frame for the $\ell$-th level, shown as output of binarization blocks of Figure 3.11, and $\gamma_B$ is an appropriate binarization threshold to be heuristically selected.[4]

---

[4] According to (3.18), the parameters $\{\alpha_\ell\}_{\ell=0}^{L-1}$ and $\gamma_B$ can be scaled by a common factor with no effect on the binarized frames. This lability is overcome by setting $\alpha_0 = 1$.



(a)

(b)

(c)

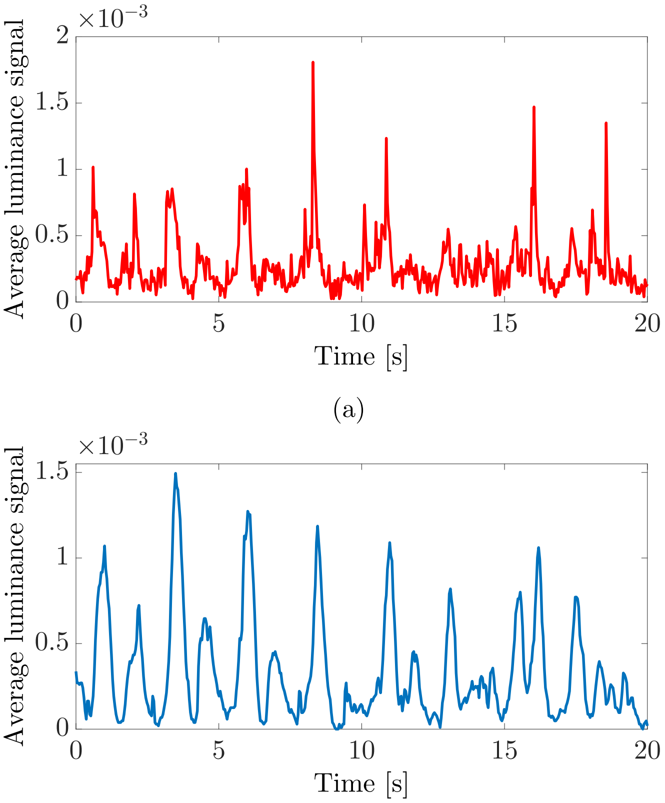
**Figure 3.13:** An illustrative example of the image processing with different temporal filters: (a) original video frames, (b) DoF filtered and binarized frames and (c) optimized IIR filtered and binarized frames with $\gamma_{\text{th}} = 15$.

Following Subsection 3.2.1, the average luminance signal for each level is computed as

$$\bar{L}_\ell[n] = \frac{1}{M_{1,\ell} M_{2,\ell}} \sum_{m_1=0}^{M_{1,\ell}-1} \sum_{m_2=0}^{M_{2,\ell}-1} B_\ell[m_1, m_2, n] \tag{3.19}$$

where $\{M_{1,\ell}\}_{\ell=0}^{L-1}$ and $\{M_{2,\ell}\}_{\ell=0}^{L-1}$ are width and height of the frames relative to each pyramidal level.

In Figure 3.13, an illustrative comparison between binarized DoF and optimized IIR filtering, without spatial decomposition and amplification, is shown, considering a few illustrative frames. More precisely, the frames are extracted from a video, taken with a smartphone, focusing on a seated breathing adult;



(a)

(b)

**Figure 3.14:** An illustrative example of extracted average luminance signal with (a) DoF filter and (b) breath-optimized IIR filter.

it can be noticed that frames shown in Figure 3.13(c) reveal better breathing movements than the ones shown in Figure 3.13(b). In this figure and in the following ones, $\gamma_B$ is set to 15.

In Figure 3.14, illustrative examples of luminance signals extracted with (a) DoF filter and (b) proposed IIR filter, are shown. In both cases, the two signals are extracted from a video framing a sleeping newborn: it can be immediately noticed that in Figure 3.14(b) the signal pattern is smoother, noise is reduced and periodicity is enhanced.

### 3.3.2 Respiratory Rate Estimation

As any level of the multi-scale decomposition shows different details, motion signals extracted from the various levels may have different features. Once motion signals are extracted, a method to estimate the RR is necessary. Since breathing fundamentally involves a quasi-periodic movement, the main purpose is to estimate the fundamental periodic component. An approach similar to the one presented in Section 3.2 is employed and extended with data fusion principles which take into account the extracted multiple motion signals.

Assuming that the motion signals represent the breathing motion at different visual details, they can be modeled, at every level $\ell = 0, 1, \ldots, L-1$, as follows:

$$\bar{L}_\ell[n] = c_\ell + A_\ell \cos\left[2\pi f_0 T_s n + \phi_\ell\right] + w_\ell[n] \tag{3.20}$$

where $c_\ell$ are continuous components, $\{w_\ell[n]\}$ are sequences of i.i.d. zero-mean Gaussian noise samples, $A_\ell$ are amplitudes, $f_0$ is the frequency of the periodic movements (common to all pyramidal levels), $\phi_\ell$ are phases, $f_s$ is sampling frequency and $n$ is the frame index. The goal is to estimate the RR by estimating $f_0$ using the various average luminance signals coming from the processed pyramidal levels.

Following the approach used for multiple camera seizure detection in [29, 30], the estimator aggregates multiple signals to reinforce the estimation of the periodic component. In order to estimate periodicity from the signals in (3.20), it is possible to apply the ML criterion. Following [46], given a vector parameter $\boldsymbol{\theta}_\ell = [A_\ell, f_0, \phi_\ell]$, relative to the $\ell$-th level, for a single signal, the strategy to estimate $\boldsymbol{\theta}_\ell$ consists of the minimization of the likelihood function

$$J_\ell\left(\boldsymbol{\theta}_\ell\right) = \sum_{n=0}^{N-1} \left(\bar{L}_\ell[n] - A_\ell \cos\left[2\pi f_0 T_s n + \phi_\ell\right]\right)^2 \tag{3.21}$$

where $N$ denotes the number of observation samples. After standard algebraic manipulations and proper simplifications, the approximate ML frequency estimator on a single motion signal becomes

$$\begin{aligned}\hat{f}_{0,\ell} &= \underset{f\in\mathcal{F}}{\operatorname{argmax}} \left| \sum_{n=0}^{N-1} \bar{L}_\ell[n] e^{-j2\pi f T_s n} \right|^2 \\ &\simeq \frac{f_s}{N} \underset{k\in\{0,1,\dots,N-1\}}{\operatorname{argmax}} \left| \mathrm{DFT}\left\{\bar{L}_\ell[n]\right\}\right|^2\end{aligned} \tag{3.22}$$

where, under the assumption of normal (periodic) breathing, the argmax search is limited to the set $\mathcal{F} = [0.25, 1.0]$ Hz, $\{0, 1, \dots, N-1\}$ is the set of discretized frequencies, $\mathrm{DFT}\{\cdot\}$ denotes the DFT operator and $\hat{f}_{0,\ell}$ is the estimate extracted on the $\ell$-th motion signal. As clear in (3.21) and (3.22), the signals $\left\{\bar{L}_\ell[n]\right\}_{\ell=0}^{L-1}$ are analyzed on sequences of $N$ temporal samples; hence the algorithm works on temporal windows of length equal to $N/f_s$ seconds.[5]

Data fusion of the $L$ signals extracted from the processed pyramids can be performed extending the ML estimator to multiple observations. In [23, 29, 30], data aggregation is used for multiple signals coming from different cameras and sensors. Using an equivalent strategy and defining $\boldsymbol{\theta} = [\mathbf{A}, f_0, \boldsymbol{\phi}]$ as the vector of the aggregated parameters, where $\mathbf{A} = [A_0\, A_1\, \dots\, A_{L-1}]$ and $\boldsymbol{\phi} = [\phi_0\, \phi_1\, \dots\, \phi_{L-1}]$, the following likelihood function can be obtained:

$$J(\boldsymbol{\theta}) = \sum_{\ell=0}^{L-1} J_\ell(\boldsymbol{\theta}_\ell) \tag{3.23}$$

where $L$ is the number of levels of the pyramid and $J_\ell(\boldsymbol{\theta}_\ell)$ are the likelihood functions for each level as defined in (3.21). Employing the different processed levels of the pyramid and the fusion technique in (3.23), the estimation of the common fundamental frequency can be obtained by the same derivation used to obtain (3.22) as:

$$\hat{f}_0 \simeq \frac{f_s}{N} \underset{k\in\{0,1,\dots,N-1\}}{\operatorname{argmax}} \sum_{\ell=0}^{L-1} \left| \mathrm{DFT}\left\{\bar{L}_\ell[n]\right\}\right|^2 \tag{3.24}$$

---

[5] To increase the estimation accuracy of the fundamental frequency, the DFT can be computed with an increased number of points $N_{\mathrm{DFT}} > N$, in order to obtain a finer set of discretized frequencies.

where $\{0, 1, \ldots, N-1\}$ is the research set of discretized frequencies. The estimated frequency $\hat{f}_0$ is declared as the RR of the monitored patient.

### 3.3.3 Performance Analysis

The system previously introduced is compared with a gold-standard polysomnographic system. In particular, the signal extracted by the pneumogram, an elastic belt that measures the changes in volume of the thoracic cavity, is used as a reference. For performance analysis, two videos of a normally breathing newborn are used (i.e., without any particular respiratory disease). The videos, which frame a newborn laterally, have resolution equal to 360 × 288 pixels and duration of, respectively, 3 min and 4 s, and 5 min and 12 s, for a total duration of 8 min and 16 s. The analysis was carried out using overall 47 interlaced windows, each lasting 20 s, with 50% overlap between consecutive windows. The system was set up with $L = 6$ levels of the Laplacian pyramid (numbered as $\ell = 0, 1, \ldots, 5$) and to employ $N/f_s = 20$ s as length of windows for the analysis of breathing movements, with $N = 500$ frames and $f_s = 25$ Hz. The cutoff frequencies of the band-pass IIR filter were set up with $f_{c,L} = 0.25$ Hz and $f_{c,H} = 1.05$ Hz and the amplification factors of the Laplacian levels were heuristically chosen as $\alpha_0 = 1$, $\alpha_1 = 3.75$, $\alpha_2 = 6.5$, $\alpha_3 = 9.25$, $\alpha_4 = 12$ and $\alpha_5 = 12$. The binarization threshold for the extraction of average luminance signals $\bar{L}_\ell[n]$ was set to $\gamma_B = 15$.

In Figure 3.15, a 20 s portion of the signal extracted from the pneumogram is compared with the signal extracted from the second ($\ell = 1$) level by our system. Every period of the quasi-periodic signal from the pneumogram corresponds to a complete respiratory act of the patient. The average luminance signal extracted by the video-processing system shows a good match with the one given by the pneumogram, considering that each respiratory act is composed of two main movements: inhalation and exhalation.

In Figure 3.16, the average luminance signals extracted from filtered and amplified pyramidal levels of a video decomposed with $L = 6$ are shown. As it can be noticed, even though the signals extracted from different levels are

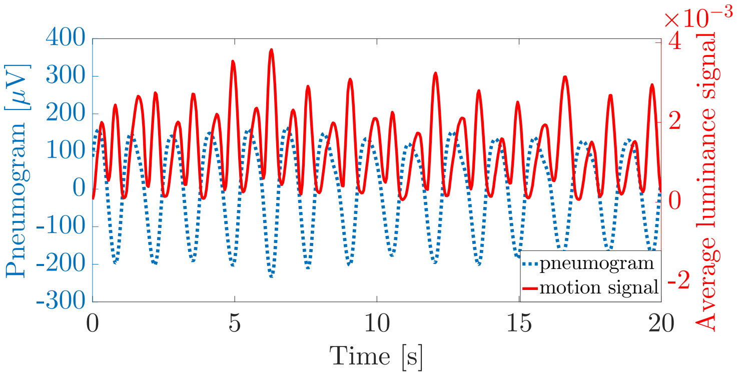


**Figure 3.15:** Comparison between the motion signal extracted by the proposed technique for the second level ($\ell = 1$) of the processed pyramid and the signal extracted from the pneumogram.

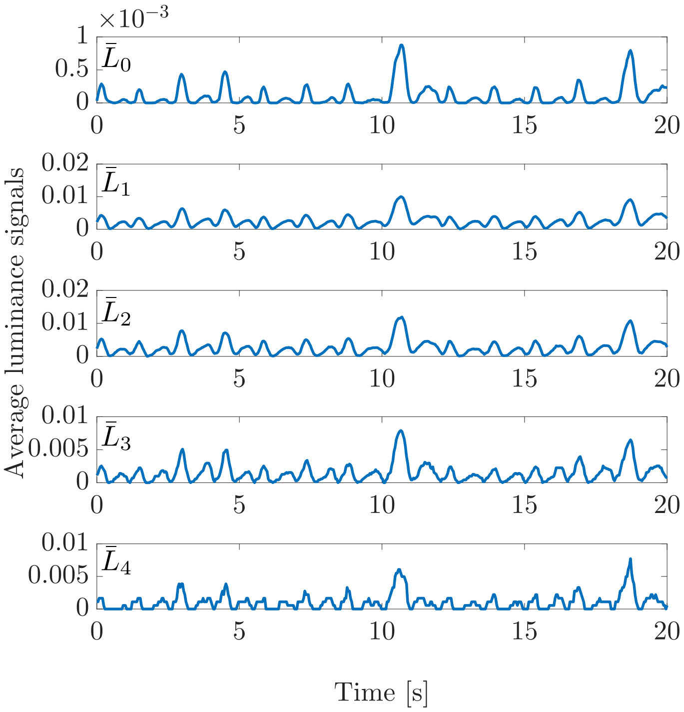
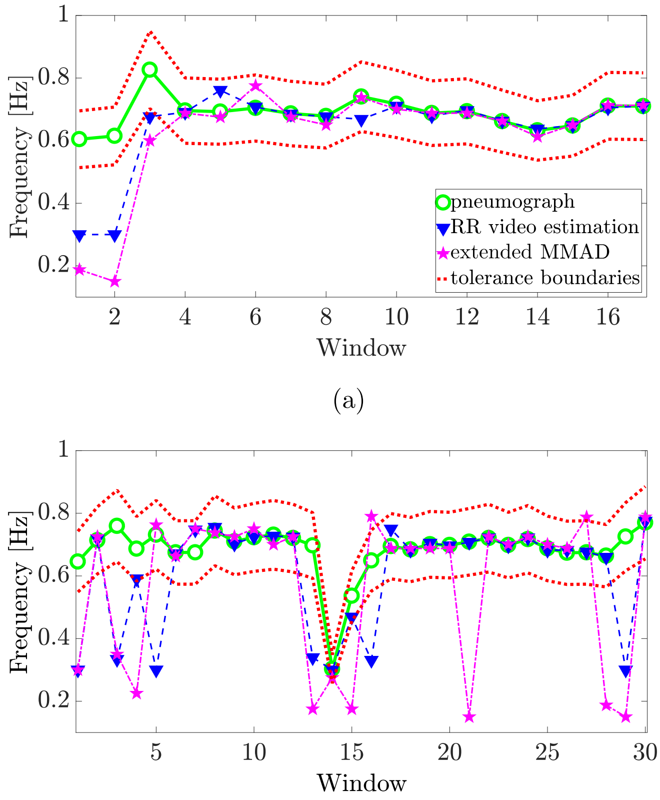


**Figure 3.16:** Average normalized luminance signals (3.19) extracted from the filtered and amplified levels of the pyramid, from the lower one to the higher one.

similar but not equal, they have all the same common frequency. The last level of the pyramid, the one corresponding to $\ell = 5$, has not been shown, because, with the resulting low detail level, a useful signal was almost absent. The breathing signal predominates in the third level ($\ell = 2$), corresponding to the motion signal $\bar{L}_2$, with a maximum normalized amplitude equal to $11.94 \cdot 10^{-3}$.

In Figure 3.17, the frequency values indicated by the pneumogram are compared with those estimated by our system—parts (a) and (b) are relative to the considered first and second video samples, respectively. In both the considered videos, the first window was discarded because of the IIR filter initial transient period, that invalidates the result. According to medical practice, we have considered ±15% as an acceptable tolerance in the estimation of the RR. The proposed algorithm is also compared with the one described in Section 3.2, here referred to as Motion Magnification for Apnea Detection (MMAD), properly extended for reporting also the estimation of the RR. In Figure 3.17(a), the RR is correctly estimated in 14 out of 16 windows in the first video sample; the same performance are obtained with the extended MMAD, as the patient is almost perfectly still. In Figure 3.17(b), the RR is correctly estimated in 24 out of 29 windows, without taking into account a movement of the newborn in the middle of the second video sample, between the 13-th and the 14-th windows, instead of 20 out of 29 windows obtained with the extended MMAD. In summary, as shown in Figure 3.17, the RR is correctly estimated in 38 out of 45 windows in the two video samples. Processing the same video samples with the extended MMAD, the RR is correctly estimated in only 34 out of 45 windows.

As an additional performance metric, the Root Mean Squared Error (RMSE) between the RR estimated by video processing and that obtained from the pneumograph, normalized to the Root Mean Squared (RMS) value of the latter, is considered. For brevity, this normalized RMSE will be also addressed as average relative error. Taking into account both video samples, the extended MMAD yields an average relative error of 30.83%. On the other hand, the algorithm here proposed yields an average relative error of 22.62%,



(a)

(b)

**Figure 3.17:** Comparison between the frequencies measured by the pneumograph, those estimated by the extended MMAD and by the here proposed method: (a) from a video sample lasting 3 min and 4 s, (b) from a video sample lasting 5 min and 12 s.

with an appreciable increase in the estimation accuracy.

## 3.4 Maximum Likelihood Video Processing

The third proposed approach does not make use of motion amplification systems; instead introduces a generalized model of pixel-wise periodicity and applies directly the ML criterion. The system first identifies the areas of the

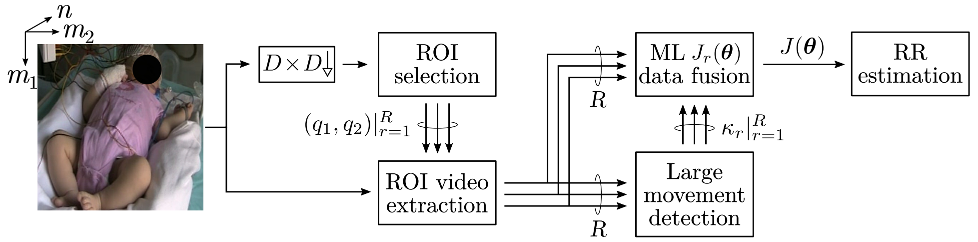


**Figure 3.18:** Block diagram of the proposed system including: ROI selection, motion detection and ML estimation of the RR.

video frames associated with breathing movements. Then, the RR is estimated by analyzing spatio-temporal intensity variations in the identified areas, without the extraction of any explicit motion signal and avoiding the interference of large movements of the patient or the medical staff.

Human breathing involves periodic movements of specific body parts and is composed of two distinct phases: inspiration and expiration. The respiratory muscles are responsible for the changes in the shape and volume of the chest cavity that cause the air movements during breathing [50]. These shape and volume changes, described by three-dimensional movements, correspond to periodic spatio-temporal small variations of the pixel intensity values in the video signal captured by a camera sensor. However, respiration generally involves subtle movements, which are difficult to detect, especially if the monitored patients are newborns. This problem increases if motion analysis algorithms process the entire framed area, which include a global view of the patient and generic environmental variations. Furthermore, framed patients may also perform large random movements, which can compromise breathing-related motion detection.

To reduce such negative effects, the proposed algorithm performs a preliminary processing which automatically identifies areas where spatio-temporal variations induced by respiration are present. The detected areas will be used as Regions of Interest (ROI) for the RR estimation. In Figure 3.18, a block diagram of the proposed system is shown: the involved steps will be described in detail in the following subsections.

The whole analysis performed by the system is based on a gray scale video $\mathbf{X}[n]$ to reduce the computational complexity. Since standard RGB cameras are used for video recordings, a color space conversion to gray scale is needed as described in Subsection 3.1.1.

Before detailing the techniques applied for the ROI selection phase and the RR estimation phase, the core method of periodicity analysis, used in both phases, is introduced.

### 3.4.1 Pixel-Wise Maximum Likelihood Estimation

Assuming that the video or a specific sub-area of the frames contain pixel intensity variations that are caused solely by the breathing movements, the related periodic and noisy video signal can be modeled in discrete time as:

$$\mathbf{X}[n] = \mathbf{C} + \mathbf{A}\cos\left(2\pi f_0 T_s n + \mathbf{\Phi}\right) + \mathbf{W}[n] \tag{3.25}$$

where the matrices have all the same size $M_1 \times M_2$, $\mathbf{C}$ describes the pixel-wise continuous components, $\mathbf{A}$ is the matrix of the amplitudes, $f_0$ is the common fundamental frequency, $T_s$ is the video sampling period, $n$ is the frame index, $\mathbf{\Phi}$ is the matrix of the initial phases, $\{\mathbf{W}[n]\}$ are matrices of i.i.d. zero-mean Gaussian noise samples and the $\cos(\cdot)$ operator and the addition of a scalar to a matrix are applied element-wise. The general aim is to estimate the fundamental frequency $f_0$, strictly related with the RR, and the amplitudes $\mathbf{A}$, useful to locate areas showing breathing movements. In order to achieve the estimation of these parameters, the application of the ML approach to the model in (3.25) is a reliable solution.

With the purpose of simplifying the notation, the operator of vectorization of a matrix representing a video frame $\mathbf{X}[n]$ is introduced:

$$\begin{aligned}\mathbf{x}_\mathrm{v}[n] = \mathrm{vec}\left(\mathbf{X}[n]\right) = \big[&X[0,0,n] \ldots X[0,M_2-1,n] \\ &X[1,0,n] \ldots X[1,M_2-1,n] \ldots \\ &X[M_1-1,0,n] \ldots X[M_1-1,M_2-1,n]\big]^T\end{aligned} \tag{3.26}$$

where the column vector $\mathbf{x}_{\mathrm{v}}[n]$ has size $M_1M_2 \times 1$ and $x_{\mathrm{v}}[p,n]$ denotes the intensity value of the $p$-th element of the $n$-th vectorized frame. The inverse operator is defined as

$$\mathbf{X}[n] = \mathrm{vec}^{-1}\left(\mathbf{x}_{\mathrm{v}}[n]\right) \tag{3.27}$$

where the $n$-th frame is retrieved to the original size $M_1 \times M_2$.

The vector parameters to be estimated can now be grouped as the vector $\boldsymbol{\theta} = [\mathbf{a}_{\mathrm{v}}, f_0, \boldsymbol{\phi}_{\mathrm{v}}]$, where, using definition (3.26), $\mathbf{a}_{\mathrm{v}} = \mathrm{vec}(\mathbf{A})$ and $\boldsymbol{\phi}_{\mathrm{v}} = \mathrm{vec}(\boldsymbol{\Phi})$. Following standard methods (see, e.g., Subsection 3.3.2 and [23]), the likelihood function to be minimized to obtain the ML estimate $\hat{\boldsymbol{\theta}}$ is

$$J(\boldsymbol{\theta}) = \sum_{p=0}^{M_1M_2-1} \sum_{n=0}^{N-1} \left[x_{\mathrm{v}}[p,n] - a_{\mathrm{v}}[n]\cos\left(2\pi f_0 T_s n + \phi_{\mathrm{v}}[p]\right)\right]^2 \tag{3.28}$$

where $N$ is the number of observed frames and $x_{\mathrm{v}}[p,n]$ represents the observed video signal. To estimate all parameters in $\boldsymbol{\theta}$, the estimate $\hat{f}_0$ is first needed; then, $\hat{\mathbf{a}}_{\mathrm{v}}$ and $\hat{\boldsymbol{\phi}}_{\mathrm{v}}$ can be estimated. Following [46, Sect. 7.10] and similarly to the method presented in Subsection 3.3.2, after proper algebraic manipulation and simplification, the following approximate ML estimator can be obtained:

$$\hat{f}_0 = \underset{f_{\min} \le f \le f_{\max}}{\mathrm{argmax}} \frac{2}{N} \sum_{p=0}^{M_1M_2-1} \left| \sum_{n=0}^{N-1} x_{\mathrm{v}}[p,n]\, e^{-j2\pi f T_s n} \right|^2 \tag{3.29}$$

where, assuming regular breathing with periodic movements, the argmax search set is limited to $[f_{\min}, f_{\max}]$, where $f_{\min}$ and $f_{\max}$, the minimum and the maximum feasible RRs, respectively, need to be properly set. This estimator is approximately ML only if the real frequency $f_0$ is not near 0 or $f_s/2$ [46].

The fundamental frequency estimator can be interpreted as the maximization of the sum of periodograms of pixel intensity variations. Neglecting constant factors, which do not affect the argmax operator, (3.29) can be rewritten as

$$\hat{f}_0 = \frac{f_s}{N} \underset{k_{\min} \le k \le k_{\max}}{\mathrm{argmax}} \sum_{p=0}^{M_1M_2-1} S_p[k] \tag{3.30}$$

where $k_{\min}$ and $k_{\max}$ are the discretized frequencies nearest to $f_{\min}$ and $f_{\max}$, respectively, and $S_p\left[k\right]$ is the discrete periodogram [46,51] computed over time at the $p$-th pixel position and defined as follows:

$$S_p\left[k\right] \triangleq \left|\sum_{n=0}^{N-1} x_{\mathrm{v}}\left[p,n\right] e^{-j2\pi\frac{k}{N}n}\right|^2 . \tag{3.31}$$

Once $\hat{f}_0$ is computed, estimates of the amplitudes related to the periodic component can be obtained as follows:

$$\hat{a}_{\mathrm{v}}[p] = \frac{2}{N}\left|\sum_{n=0}^{N-1} x_{\mathrm{v}}[p,n] e^{-j2\pi\hat{f}_0 T_s n}\right| . \tag{3.32}$$

The distribution of amplitudes related to the estimated fundamental frequency for each pixel location can be retrieved, using (3.27), as $\hat{\mathbf{A}} = \mathrm{vec}^{-1}\left(\hat{\mathbf{a}}_{\mathrm{v}}\right)$.

According to (3.28)–(3.32), the video signal is analyzed on sequences of $N$ frames: hence, the algorithm works with disjoint temporal windows of length equal to $NT_s$. Furthermore, in order to obtain a finer estimation over time, interlaced windows may be used, as applied also in the methods in Sections 3.2 and 3.3.

The described algorithm can also be seen as a data fusion method where the involved signals are the temporal intensity variations of each pixel in the video signal. With reference to Figure 3.18, the ML approach will be used in the blocks which perform ROI selection, ML data fusion and RR estimation.

### 3.4.2 ROI Selection and Large Motion Management

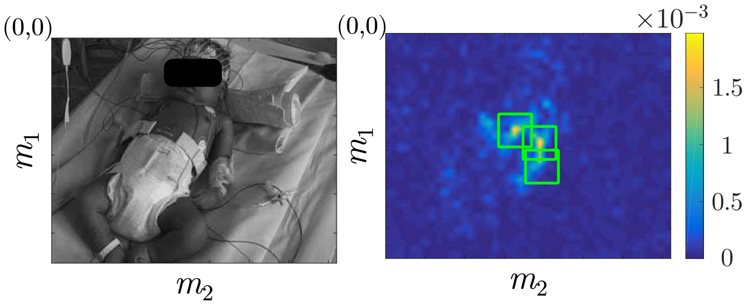
In Subsection 3.4.1, the general ML approach assumes that only quasi-periodic respiratory movements are framed from the video camera. This assumption is quite strong, considering that cameras employed for video monitoring typically frame the whole patient and not a specific body part only. The framed area can also be subject to general body movements of the patient, passive body movements due to clinical care and generic environmental changes. A possible solution, which both tries to solve these problems and limit the computational complexity of the algorithm, is to select $R$ ROI, with specified size of $W \times W$ pixels, and apply the RR estimation algorithm only to the selected areas of the video. The main problem is finding a method for automatic selection of such regions.

First of all, $Q$ frames of the video should be selected where only respiration is present. To avoid large movements or environmental changes, a motion detection algorithm, inspired by [28], is applied. Using the $n$-th temporal window of the video frame $\{\mathbf{X}\,[n]\}_{n=0}^{N-1}$, a simple Finite Impulse Response (FIR) filter is employed to compute the difference between consecutive frames as expressed in (3.2) and where the output frames are named $\mathbf{Y}\,[n]$. In order to obtain a quantitative indicator of the framed large movements, an average motion signal is computed as

$$\bar{L}[n] = \frac{1}{M_1 M_2} \sum_{m_1=0}^{M_1-1} \sum_{m_2=0}^{M_2-1} \big| Y\,[m_1, m_2, n]\,\big|. \tag{3.33}$$

If the average motion signal $\left\{\bar{L}[n]\right\}_{n=0}^{Q-1}$ shows that large movements or intensity changes are not present, the corresponding $Q$ frames are used for ROI selection.

Once $Q$ frames of the video are selected, this video block is used to detect the $R$ desired ROI. Each frame of the block $\{\mathbf{X}[n]\}_{n=0}^{Q-1}$ of size $M_1 \times M_2$ is spatially decimated by an integer $D$ over both dimensions (the $D \times D{\downarrow}$ block in Figure 3.18), in order to reduce the number of pixels to be analyzed and the overall computational complexity of the ML approach. The decimated video sequence $\{\mathbf{X}_D[n]\}_{n=0}^{Q-1}$ includes frames of size $\lceil M_1/D \rceil \times \lceil M_2/D \rceil$, where $\lceil \cdot \rceil$ denotes the smallest integer larger than or equal to the argument. Next, the ML estimation described in Subsection 3.4.1 is applied to the video sequence $\{\mathbf{X}_D[n]\}_{n=0}^{Q-1}$—the fundamental frequency $\hat{f}_0$ is estimated and associated with the RR of the framed patient. This processing leads to the pixel-wise amplitudes matrix $\hat{\mathbf{A}}_D$ related to the motion estimation in $\{\mathbf{X}_D[n]\}_{n=0}^{Q-1}$, obtained similarly to (3.32). The obtained matrix is interpolated to the original frame


[redacted]

(a) (b)

**Figure 3.19:** Examples of (a) a sample gray scale image $\mathbf{X}[n]$ of a video framing a newborn and (b) the relative estimated amplitudes matrix $\hat{\mathbf{A}}_{D,i}$ used for ROI selection and the selected regions with $R = 3$.

size of $M_1 \times M_2$ by two-dimensional bicubic interpolation [52]: the result, denoted as $\hat{\mathbf{A}}_{D,i}$, can be considered an approximation of the matrix $\hat{\mathbf{A}}$ computed over the whole original frames. In Figure 3.19, in (a) an illustrative image of the framed patient is shown and in (b) the corresponding estimated amplitudes $\hat{\mathbf{A}}_{D,i}$ for ROI detection are shown: larger values can be observed in the areas mainly involved with periodic movements of the patient. In the shown example, the selected areas are near the chest of the newborn.

Once $\hat{\mathbf{A}}_{D,i}$ is extracted, from the previously selected video block $\{\mathbf{X}[n]\}_{n=0}^{Q-1}$, the centers of the $R$ regions are selected adopting the algorithm described by the pseudo-code in Figure 3.20, where $\mathcal{R}$ and $\mathcal{M}$ are the sets of the centers of the searched ROI and the pixel coordinates of the video frames, respectively (ROI selection block in Figure 3.18). In Figure 3.19(b), the selected regions with $R = 3$ for the previously introduced matrix $\hat{\mathbf{A}}_{D,i}$, indicated by the highlighted (green) boxes, are also shown. The procedure for ROI selection is repeated periodically, in order to follow possible changes of position of the framed patient.

Once ROI with breathing movements are identified, the ML approach described in Subsection 3.4.1 is applied to each region. To reduce the effects of random movements during estimation of the fundamental frequency inside the

**Algorithm for ROI selection**

```
$\mathcal{R} \leftarrow \emptyset$
$\mathcal{M} \leftarrow \{(m_1, m_2) : \lfloor W/2 \rfloor \leq m_1 \leq M_1 - 1 - \lfloor W/2 \rfloor,$
        $\lfloor W/2 \rfloor \leq m_2 \leq M_2 - 1 - \lfloor W/2 \rfloor\}$
$r = 1$
repeat
    $(q_1, q_2) \leftarrow \text{argmax}_{(m_1,m_2)\in\mathcal{M}} \hat{\mathbf{A}}_{D,i}$
    $\mathcal{R} \leftarrow \mathcal{R} \cup \{(q_1, q_2)\}$
    $\mathcal{M} \leftarrow \mathcal{M} \setminus \{(m_1, m_2) :$
                $q_1 - \lfloor W/2 \rfloor \leq m_1 \leq q_1 + \lfloor W/2 \rfloor,$
                $q_2 - \lfloor W/2 \rfloor \leq m_2 \leq q_2 + \lfloor W/2 \rfloor\}$
    $r \leftarrow r + 1$
until $r \leq R$
```

**Figure 3.20:** Pseudo-code of the algorithm for ROI selection by extraction of the ROI center coordinates.

analyzed temporal window, the motion detection algorithm described above is first applied to each region, as shown by the motion detection block in Figure 3.18. Defining as $\{\mathbf{X}_r[n]\}_{n=0}^{N-1}$ the frames of the considered temporal window of the $r$-th extracted region ($r \in \{1, 2, \ldots, R\}$), the simple FIR filtering method in (3.2) is applied, obtaining the output $\mathbf{Y}_r[n]$. The average motion signal of the $r$-th region is computed as

$$\bar{L}_r[n] = \frac{1}{W^2} \sum_{m_1=0}^{W-1} \sum_{m_2=0}^{W-1} \left| Y_r[m_1, m_2, n] \right|. \tag{3.34}$$

These $R$ signals can be used to detect the ROI affected by undesired large movements according to the following decision rule:

$$\kappa_r = \begin{cases} 0 & \text{if } \exists n : \bar{L}_r[n] > \gamma_{\text{th}} \\ 1 & \text{else} \end{cases} \qquad r = 1, 2, \ldots, R \tag{3.35}$$

where $\kappa_r$ is a binary parameter defining the presence/absence of large motion inside the $r$-th ROI and $\gamma_{\text{th}}$ is a heuristically chosen threshold for the detection

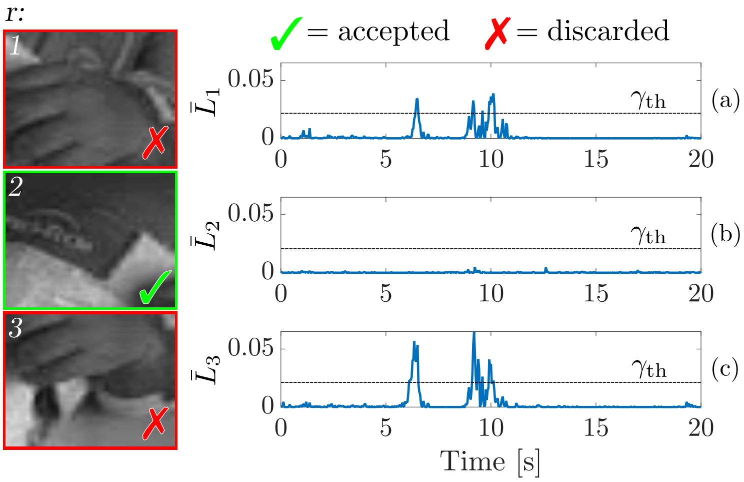


**Figure 3.21:** Extracted motion signals $L_r$ for the $R = 3$ ROI. For $r = 1, 3$, the regions are discarded due to large body movements.

of large values in $\bar{L}_r[n]$. In Figure 3.21, the extracted motion signals for the previously selected ROI of Figure 3.19(b) are shown: the first an the third regions are discarded, setting $\kappa_1 = \kappa_3 = 0$ because of a large motion of the left arm of the newborn described by signals $\bar{L}_1$ and $\bar{L}_3$, shown in Figure 3.21(a) and Figure 3.21(c), respectively. In Figure 3.21(b), the signal for the second region is shown: it stays below the motion threshold $\gamma_{\text{th}}$, so this region is considered valid ($\kappa_2 = 1$) for RR extraction.

### 3.4.3 Respiratory Rate Estimation

In the last step, following the information fusion introduced in Subsection 3.4.1 and defining as $J_r(\boldsymbol{\theta})$ a likelihood function similar to (3.28) but referred to the $r$-th region ($r \in \{1, 2, \ldots, R\}$), the RR in the analyzed temporal window is obtained by using the ML technique described in (3.28)–(3.32) employing the following likelihood function relative to the active ROI:

$$J(\boldsymbol{\theta}) = \sum_{r=1}^{R} \kappa_r J_r(\boldsymbol{\theta}) . \tag{3.36}$$

The fundamental frequency estimated by maximizing (3.36) is defined as the RR of the framed patient in the analyzed temporal window, as shown in the final step in Figure 3.18.

### 3.4.4 Performance Analysis

In this section, the RR estimation algorithm is compared with data recorded from the pneumogram, one of the sensors of the gold standard polysomnographic device. Moreover, a performance comparison with the video processing-based algorithm for RR estimation described in Section 3.3, based on subtle periodic motion analysis and here referred to as spatio-temporal video-processing for RR estimation (STRE), is carried out.

The proposed algorithm is tested on video streams recorded in the NICU of the University Hospital of Parma, with a resolution of $360 \times 288$ pixels and a sampling rate of $f_s = 25$ Hz. Temporal windows of length $NT_s = 20$ s, with a 50% interlacing factor, are used for the analysis. The range of interest for the fundamental frequency search, defined according to the feasible RR of the newborn, is set as $f_{\min} = 0.3$ Hz and $f_{\max} = 1.2$ Hz, corresponding to 18 and 72 breaths/min, respectively. The number $R$ of ROI is set to 4 with corresponding size of $20 \times 20$ pixels ($W = 20$). For the selection of such regions, a downsampling factor $D = 8$ is applied and $Q = 500$ frames are used. The threshold for large motion detection is chosen depending on the camera viewpoint; its value is typically selected in the range $\gamma_{\text{th}} \in [0.0175, 0.1]$ for pixel intensities in the set $\{0, 1, \ldots, 255\}$.

Pixel intensity variations in specific locations are directly connected with the spatio-temporal movements of respiration: in Figure 3.22, an illustrative camera view of a newborn patient is shown. The area near the throat of the framed newborn is zoomed in and a short row of 6 pixels, with global coordinates, referring to the left image of Figure 3.22, with $m_1 = 230$ and $m_2 \in \{215, 216, \ldots, 220\}$, is highlighted. Inside these pixels, a quasi-periodic movement from left to right of the throat is expected: temporal intensity variations in this area are now analyzed. In Figure 3.23(a), examples of gray scale

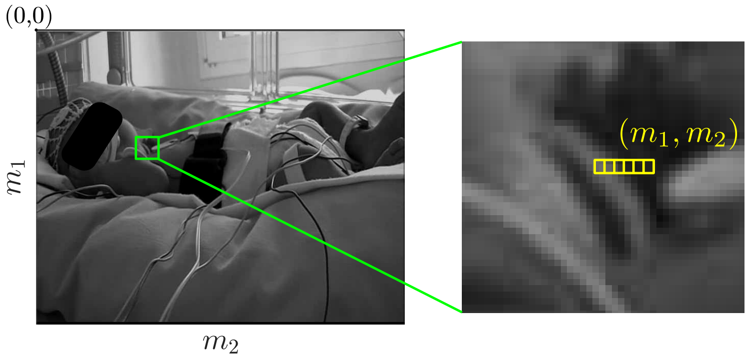
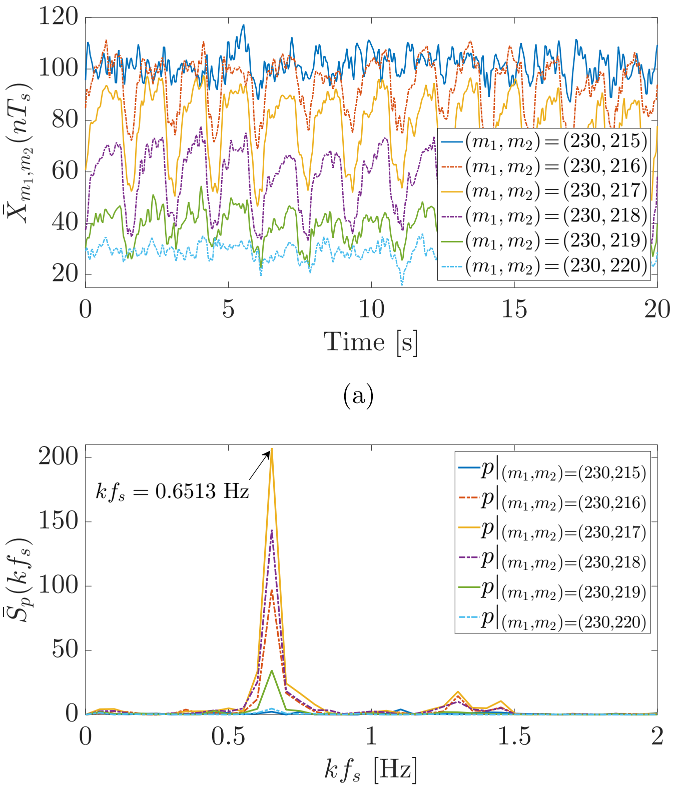


**Figure 3.22:** An illustrative example of a framed regularly breathing patient, with sensors of the polysomnographic system applied. Zoom in the area near the throat with a selected row of pixels for the analysis of intensity variations is also shown (yellow box).

pixel intensities $\bar{X}_{m_1,m_2}(nT_s) = X[m_1, m_2, n]$ of the region selected in Figure 3.22 on a temporal window of 20 s, with a framed regularly breathing newborn at approximately 0.66 Hz, are shown. As it can be noticed, the pixels at the far ends of the area are not affected by significant periodic variations, but the respiratory motion is clearly visible in the intensity changes of the central pixels of the sample row. Pixels with similar variations can be localized by the presented ML approach and used with the described fusion technique to reinforce RR estimation using regions on different body parts. In fact, a similar behavior can be identified in an area near the chest of the newborn. In Figure 3.23(b), the periodograms $\bar{S}_p(kf_s) = S_p[k]$ of the signals graphed in Figure 3.23(a) are shown: in order to better visualize the frequencies of interest, the continuous component of each considered signal $X[m_1, m_2, n]$ is removed. As expected, the signals with low periodic component have smaller contribution at the frequencies related to respiration. The estimated RR in this analyzed temporal window is $\hat{f}_0 = 0.6513$ Hz, which is very close to the value of the real RR of the patient.

For performance evaluation, two video recordings (framed laterally) of a newborn, breathing normally and without any respiratory disorder, have been



(a)

(b)

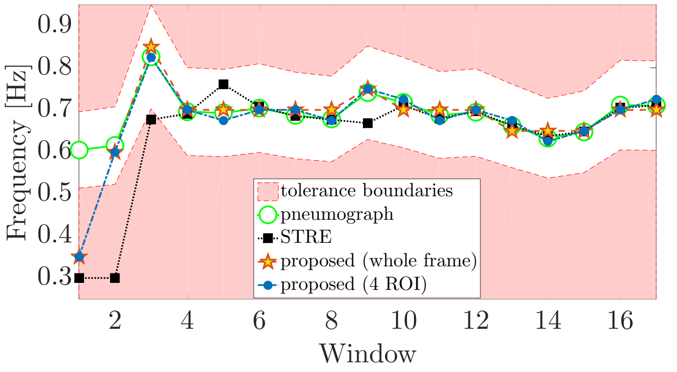
**Figure 3.23:** An example of the analysis on a selected temporal window of a newborn regularly breathing: (a) the pixel-wise intensity variations over time and (b) a frequency analysis of the corresponding extracted signals.

employed. The selected videos have time durations of 3 min and 4 s and of 5 min and 12 s, respectively. The total duration of the videos processed to test the algorithm is thus 8 min and 16 s. These video samples are the same ones employed for performance evaluation in Subsection 3.3.3.

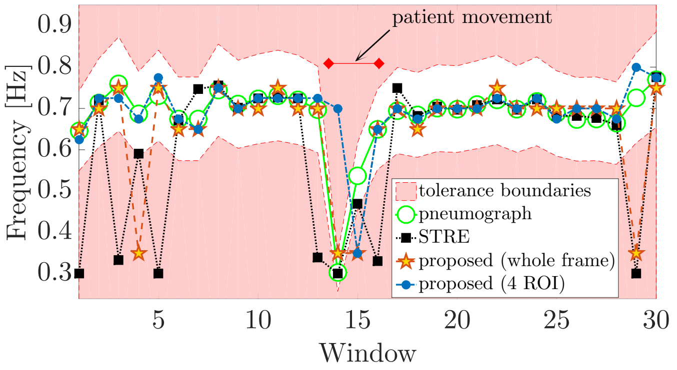
In Figure 3.24, the values of RRs indicated by the pneumogram device are compared with those estimated by the proposed system. The algorithm is tested on the whole video frames (without ROI selection) and using $R = 4$

ROI, to show the usefulness of large motion detection and removal. In both the considered tests, the first window must not be considered because of incomplete data acquired due to the interlaced window analysis. A RR estimation tolerance of ±15%, with respect to the pneumographic reference, is assumed. In Figure 3.24, invalid rates, with respect to those indicated by the considered gold standard, are presented in the shaded areas. In Figure 3.24(a), results relative to the first video are illustrated: the RR is correctly estimated by the proposed algorithm in all the 16 useful windows. The method obtains similar results when applied to either the whole frame or the 4 ROI, because the newborn patient is still. In Figure 3.24(b), results concerning the second video are shown. Here, the RR is correctly estimated in 25 out of 29 useful windows, by using the entire frame, and in 27 out of 29 useful windows, by using 4 ROI. These results are obtained only excluding the first analyzed window but without taking into account a movement of the newborn in the middle of the video, shown in Figure 3.24(b) and concentrated approximately between the 13-th and the 14-th temporal windows. Here, some large random movements of the newborn are recorded: the application of the ROI method for large motion avoidance clearly improves the algorithm performance. In both the employed video tests, the RR is correctly estimated, overall, in 43 out of the 45 considered windows. Comparing the current results with those obtained for the STRE algorithm in Section 3.3, a performance improvement can be noticed. In fact, the number of windows in which the RR is correctly estimated increases: from 14 (with STRE) to 16 (with the proposed algorithm), for the first considered video sample, and from 24 to 27 for the second video sample. In the latter, newborn movements, which involve the whole body, alter also the pneumograph signal, so that a finer estimation of the RR is not feasible.

A performance evaluation of the algorithms in terms of RMSE, as previously introduced in Subsection 3.3.3, is also reported. An overall performance evaluation about the estimation of the RR on the two tested video samples is reported. Table 3.2 shows performance of the three proposed methods for RR estimation by video processing discussed in this chapter in terms of RMSE,



(a)



(b)

**Figure 3.24:** Comparison between RRs estimated by the pneumograph, the algorithm introduced in Section 3.3 and the proposed method: (a) first video sample (duration: 3 min and 4 s) and (b) second video sample (duration: 5 min and 12 s).

specified in Hertz (Hz), and average relative error, computed as normalization of the RMSE with respect to the RMS value of the real RR, clearly proving the increase in estimation accuracy of the fundamental frequency related to the RR of the framed patient. The algorithm proposed in this section is here referred to as Maximum Likelihood Video Processing for RR estimation (MLVP). Taking into account both video samples and normalizing the RMSE with respect

| **Algorithm** | **RMSE** [Hz] | **average relative error** |
|---|---|---|
| Extended MMAD [Section 3.2] | 0.218 | 30.83% |
| STRE [Section 3.3] | 0.156 | 22.62% |
| MLVP (whole frame) [Section 3.4] | 0.089 | 12.84% |
| MLVP (4 ROI) [Section 3.4] | 0.076 | 11.03% |

**Table 3.2:** Performance comparison in terms of RR estimation error for the three proposed algorithms.

to the RMS value of the frequencies estimated by the pneumograph, the here proposed algorithm obtains an average relative error of 12.84% and 11.03% analyzing the entire frame and the 4 ROI method, respectively. The increase in accuracy on RR estimation with respect to the STRE algorithm, reporting an average relative error equal to 22.62%, is clear.

As a third test, a simulated video obtained as discussed in the following Chapters 4 and 5 is considered (see also [53]). Since the time duration of this video, 16 min and 58 s with a window interlacing factor of 95%, is longer than the previous ones, with a total number of 1017 windows, more concise performance metrics are now considered. Table 3.3 reports the total number of windows in which RR is correctly detected and the RMSE between the estimated and simulated RRs, normalized to the RMS value of the latter. The obtained results clearly indicate that the MLVP algorithm exhibits a higher

| **Algorithm** | **NW** | **CD** | **average relative error** |
|---|---|---|---|
| STRE | 1017 | 895 | 16.97% |
| MLVP (4 ROI) | 1017 | 955 | 13.71% |

Legend: NW = Number of useful Windows; CD = windows in which RR is Correctly Detected.

**Table 3.3:** Performance in terms of RR estimation tested on simulated video samples.

robustness to large body movements and a higher RR estimation accuracy with respect to the STRE algorithm.

## 3.5 Conclusion

In this chapter, three different methods for the detection of apnea and estimation of the RR have been presented and tested on newborn patients.

The first method makes use of a previously developed algorithm for the analysis of movements recorded by video cameras and solves the problem of revealing subtle breathing movements, otherwise invisible, employing a renowned motion magnification system. Results demonstrate that the method has been able to detect severe apnea events in infants, in accordance with the clinical examination.

The second method overcomes some issues of the procedure for the extraction of motion information related to periodic movements and integrates it with the motion magnification algorithm. The ability of this method to clearly extract breathing signals is discussed and the performance on RR estimation is analyzed.

Finally, a third method for RR estimation is presented, which avoids the magnification system and analyzes pixel-wise temporal variations by a generalized ML approach. Then, a performance analysis on RR is presented, demonstrating the improvement with respect to the previous algorithms.

# Part II

# Models and Simulators of Respiration

# Chapter 4

# Statistical Models of Breathing Patterns

## 4.1 Introduction

A difficulty in the design of video processing-based algorithms is the lack of large databases of relevant video recordings obtained from real patients with respiratory disorders and properly matched with reliable medical data. This may be due, for example, to the rarity of some diseases, such as Congenital Central Hypoventilation Syndrome (CCHS) [8] and severe apnea events, especially in full-term newborns. This makes the design and optimization of video processing-based algorithms quite critical. For this reason, the development of a statistical model of Respiratory Rate (RR) patterns, including the occurrence of respiratory pauses and apnea events, may be of significant interest. Such a model can be very useful in order to devise realistic simulators, as discussed in Chapter 5, and create a large set of video recordings which allow a more efficient design and reliable test of automatic video processing-based RR estimation and apnea detection systems.

In this chapter, statistical models, based on a Continuous-Time Markov Chain (CTMC), aimed at describing the main features of breathing and of respiratory diseases (e.g. apneas) and realistic RR patterns derived from medical data, are presented. The parameters of both models are extracted by inference systems for continuous-time Markov random processes. Finally, the statistical models are validated, comparing results of simulations of the CTMCs with data extracted from real cases.

The chapter is organized as follows. In Section 4.2 a simple two-state CTMC model to represent the behavior of respiratory pauses and apnea episodes is presented. In Section 4.3 the previous statistical model is extended to a multiple-state CTMC in order to represent also complete breathing patterns of patients with regular respiration or affected by apnea events. The validation of the statistical models is addressed in Section 4.4. Finally, in Section 4.5, final remarks are given.

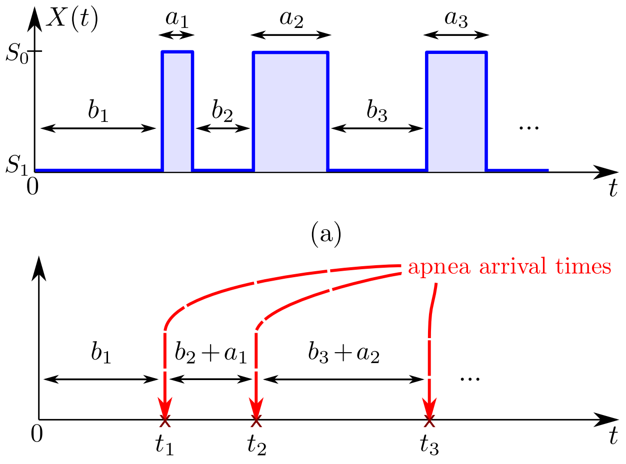
## 4.2 Model for Apnea Episodes

Given the inability to forecast the starting time and duration of an apnea event, it may be useful to introduce a probabilistic model allowing to describe and approximately replicate the main features of such events in an appropriate stochastic model. Apneas could be simply defined as events causing complete respiratory interruption or significant slowdown, during which the normal breathing pattern is modified. Hence, it is possible to describe the overall breathing pattern of a patient affected by apneas, in a simplified manner, by means of alternating intervals with normal breathing and apnea events.

In order to model the behavior of apneas, it is useful to introduce a simple two-state system, where the states $\{S_0, S_1\}$ are defined as follows:

$$\begin{aligned} S_0 &\triangleq \{\text{apnea}\} \\ S_1 &\triangleq \{\text{regular breathing}\}\,. \end{aligned} \tag{4.1}$$

By construction, at a generic time $t$, a patient can only be in one of the two states in (4.1) and the state process is denoted as $X(t)$. The durations of



(a)

(b)

**Figure 4.1:** An example of arrival process of apneas, showing (a) durations of apnea events and (b) inter-arrival times between consecutive starting epochs.

regular breathing intervals and apnea events, namely sojourn times in states $S_1$ and $S_0$, respectively, are the characteristic statistical quantities of the proposed model. To derive a stochastic model, it is useful to define the following sequences of random variables:

$$\begin{aligned} b_i &\triangleq \{\text{time duration of normal breathing between} \\ &\quad\;\; \text{the } (i-1)\text{-st and the } i\text{-th apnea event}\} \\ a_i &\triangleq \{\text{time duration of the } i\text{-th apnea event}\} \end{aligned} \tag{4.2}$$

where $i \in \mathbb{N}$ denotes the index of the regular breathing phase or apnea episode. It can be assumed that each of the two sequences of random variables $\{b_i\}$ and $\{a_i\}$ are independent and identically distributed (i.i.d.), since the durations of disjoint apneas are independent of each other and the arrival times do not depend on previous events. In Figure 4.1(a), a graphical example of the respiratory process $X(t)$ is shown: the occurrences of apneas are highlighted in Figure 4.1(b) with corresponding arrival times. Global arrival times of apneas

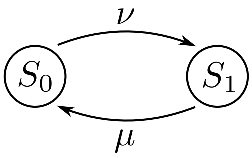


**Figure 4.2:** Two-state CTMC state diagram associated with the model for apnea episodes.

can be expressed as

$$t_i = b_i + \sum_{j=1}^{i-1} (b_j + a_j) \,. \tag{4.3}$$

In order to make the proposed model analytically tractable, the random variables introduced in (4.2) are assumed to have the following distributions:

$$\begin{aligned} b_i &\sim \exp(\mu) \\ a_i &\sim \exp(\nu) \end{aligned} \tag{4.4}$$

where the notation $\sim$ is used to specify the distribution and $\exp(\cdot)$ denotes the exponential distribution [54] with parameter $\mu$ or $\nu$, respectively—these parameters have to be properly chosen to describe realistic durations of regular breathing periods and apneas. Owing to the previous assumptions, the system model can be described by a CTMC [55] with two states. It thus follows that the sojourn time in each state is an exponentially distributed random variable and the process $X(t)$ exhibits the memoryless property [54, 56].

The introduced CTMC model has the state diagram shown in Figure 4.2: by construction, the transition rate from state $S_1$ to state $S_0$ is $\mu$, whereas the transition rate from state $S_0$ to state $S_1$ is $\nu$. Defining $\lambda_{m,n}$ as the transition rate from state $m$ to state $n$, $m, n \in \{0, 1\}$, the complete transition matrix [54] for the CTMC shown in Figure 4.2 is:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{0,0} & \lambda_{0,1} \\ \lambda_{1,0} & \lambda_{1,1} \end{bmatrix} = \begin{bmatrix} -\nu & \nu \\ \mu & -\mu \end{bmatrix} . \tag{4.5}$$

Defining the probability of being in a generic state $S_j$ at generic time $t$ as

$p_j(t) = P\{X(t) = j\}$, with $j \in \{0, 1\}$, it is possible to define the state probability vector at a generic time $t$ as

$$\mathbf{p}(t) = \begin{bmatrix} p_0(t) \\ p_1(t) \end{bmatrix}. \tag{4.6}$$

The stationary distribution $\boldsymbol{\pi}$ of the CTMC [54], defined as

$$\boldsymbol{\pi} = \lim_{t \to +\infty} \mathbf{p}(t) \tag{4.7}$$

is given by the following vector

$$\boldsymbol{\pi} = \begin{bmatrix} \frac{\mu}{\mu+\nu} \\ \frac{\nu}{\mu+\nu} \end{bmatrix}. \tag{4.8}$$

Setting accurately the average values of the durations of regular breathing and apnea events, given, respectively, by

$$\begin{aligned} \mathbb{E}\{b_i\} &= \tfrac{1}{\mu} \\ \mathbb{E}\{a_i\} &= \tfrac{1}{\nu} \end{aligned} \tag{4.9}$$

it is possible to use the CTMC-based statistical model of the respiratory process described above to generate the desired values $\{b_i\}$ and $\{a_i\}$ for the sojourn times in state $S_1$ and in state $S_0$, respectively. The pseudo-randomly generated sequences $\{b_i\}$ and $\{a_i\}$ can be used by a properly designed simulator to replicate a behavior similar to real apnea events, as discussed in Chapter 5. The average values in (4.9), can be estimated measuring the mean duration of regular breathing periods and apnea/respiratory pauses in monitored patients, respectively.

The following Section 4.3 extents the here presented model to a generic $N$-state one, able to represent also complete breathing patterns and the variations over time of the RR.

## 4.3 Model of Breathing Patterns

The two-state CTMC described in the previous section can be extended to describe more complex breathing patterns. In fact, the RR changes over time, depending on the patient physical activity and health conditions. Normally, the RR of a patient at rest is age-dependent and typically ranges from 30 bpm to 60 bpm (equivalent to 0.5–1.0 Hz) for newborns and from 12 bpm to 20 bpm for adults (equivalent to 0.2–0.333 Hz) [57].

### 4.3.1 Extended Markov Chain Model

In order to devise a simple model for a complete RR pattern, it is useful to introduce a finite set of states $\mathcal{S} = \{S_0, S_1, \ldots, S_{N-1}\}$. State $S_n$, with $n \in \{0, 1, \ldots, N-1\}$, describes breathing with a RR denoted as $\varrho_n \geq 0$. Occurrence of respiratory pauses or apnea events and large random movements of the patient body are also considered. The statistical model of the RR pattern can encompass all the following conditions.

- If the patient is regularly breathing, i.e. he/she is not suffering from apnea events and no large random body movements appear, the states $\{S_0, S_1, \ldots, S_{N-1}\}$ are used to describe regular RRs, characterized by values $\{\varrho_n\}_{n=0}^{N-1}$ with $\varrho_n \in [R_L, R_H]$, where $R_H > R_L \geq 0$ denote highest and lowest admissible RRs, respectively.
- If the patient is affected by respiratory pauses/apneas, the state $S_0$ is reserved to represent this condition, so that $\varrho_0$ is formally set to 0, to describe absence of breathing, and states $\{S_1, S_2, \ldots, S_{N-1}\}$ are considered to model regular breathing.
- If the patient is subject to large random body movements, during which the RR is undetectable, the state $S_{N-1}$ is reserved to represent this condition. The RR $\varrho_{N-1}$ is set to an arbitrary value $R_M$ much larger than the physically acceptable ones: more precisely, $\varrho_{N-1}$ is set to $R_M \gg R_H$. States $\{S_0, S_1, \ldots, S_{N-2}\}$ are used to represents regular RRs.
- If the patient is both suffering from respiratory pauses/apneas and subject to large random body movements, the states $S_0$ (with $\varrho_0 = 0$) and $S_{N-1}$ (with $\varrho_{N-1} = R_M \gg R_H$) are reserved for absence of

breathing and random movements, respectively. The remaining states $\{S_1, S_2, \ldots, S_{N-2}\}$ are used to describe regular breathing.

The following ordering is assumed: $\varrho_0 < \varrho_1 < \cdots < \varrho_{N-1}$. Since the RR is inherently continuous-valued, each state represents an approximation of the real RR. Therefore, the set $\mathcal{S}$ represents a finite state model of a discrete-valued process approximating the overall RR pattern—the larger the number $N$ of states, the better the approximation at the cost of a higher modeling complexity.

According to the above statistical model, the RR process, denoted as $X(t)$, is defined as a continuous-time process with state space $\mathcal{S}$. The time intervals during which the patient is breathing with rate $\varrho_n$ or is subject to an apnea/respiratory pause or large body movements, namely the sojourn times in the corresponding state $S_n$, can be modeled as random variables and the introduced random process $X(t)$ can be generally described as a Markov process. Ignoring the influence of other vital signs which can modify the RR of a patient over time, such as the heart rate or the oxygen saturation in the blood, the RR pattern cannot be predicted. To derive a model that approximates this stochastic behavior, the random variable $\tau_\ell$ is introduced, which specifies the $\ell$-th sojourn time, where $\ell \in \mathbb{N}^+$ is an index that counts the number of state changes. Jump times can be expressed, in terms of sojourn times, as

$$t_\ell = \sum_{q=1}^{\ell} \tau_q. \tag{4.10}$$

In Figure 4.3, a graphical example of the modeled finite-state RR process $X(t)$ is shown, with highlighted sojourn times and change of state instants.

Since the influence of other vital signs is ignored, it can be assumed that the random variables $\{\tau_\ell\}$ are independent, so that the process $X(t)$ exhibits the memoryless property [54, 56]. Accordingly, the $\ell$-th sojourn time, conditioned on state $S_n$, has the following exponential distribution:

$$\tau_\ell \sim \exp(\mu_n) \tag{4.11}$$

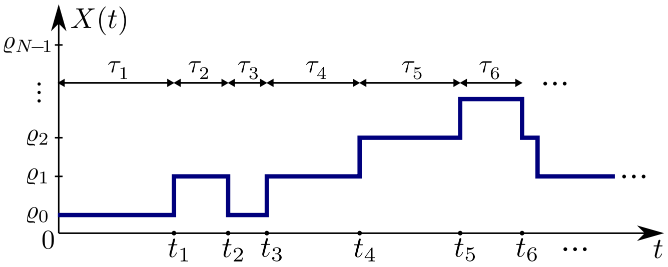


**Figure 4.3:** An example of RR pattern modeled by the finite set $\mathcal{S}$, showing sojourn times and jump times.

where $\mu_n$ is the parameter of the distribution and can be interpreted as the rate at which the process $X(t)$ leaves the state $S_n$ [56]. In Figure 4.4, a generic state diagram of the proposed CTMC is shown, where $\lambda_{m,n}$, $m, n \in \{0, 1, \ldots, N-1\}$, denotes the transition rate from state $S_m$ to state $S_n$.

The statistical behavior of a CTMC is characterized by its infinitesimal generator (or transition rate) matrix [54]

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{0,0} & \lambda_{0,1} & \cdots & \lambda_{0,N-1} \\ \lambda_{1,0} & \lambda_{1,1} & \cdots & \lambda_{1,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{N-1,0} & \lambda_{N-1,1} & \cdots & \lambda_{N-1,N-1} \end{bmatrix} \tag{4.12}$$

where the entries and the parameters in (4.11) are related as

$$\mu_m = -\lambda_{m,m} = \sum_{\substack{n=0 \\ n \neq m}}^{N-1} \lambda_{m,n}. \tag{4.13}$$

At the end of each waiting time, a state transition occurs: the arrival state is determined according to the transition probabilities of the embedded Markov chain [56, Sect. 11.4], which can be obtained by the infinitesimal generator matrix $\mathbf{\Lambda}$ in (4.12) as follows:

$$\mathbf{Q} = \mathbf{I} - [\operatorname{diag}(\mathbf{\Lambda})]^{-1} \mathbf{\Lambda} \tag{4.14}$$

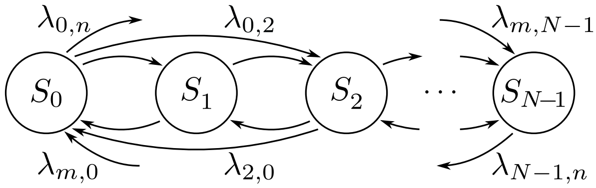


**Figure 4.4:** State diagram of the CTMC model with respective transition rates associated with the model of breathing patterns.

where $\mathbf{I}$ is the $N \times N$ identity matrix and $[\operatorname{diag}(\mathbf{\Lambda})]^{-1}$ is the inverse of the diagonal matrix

$$\operatorname{diag}(\mathbf{\Lambda}) = \begin{bmatrix} \lambda_{0,0} & 0 & \cdots & 0 \\ 0 & \lambda_{1,1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{N-1,N-1} \end{bmatrix}. \tag{4.15}$$

As the embedded Markov chain of the CTMC is assumed ergodic, the stationary distribution, described by an $N$-element vector $\boldsymbol{\pi}$, can be obtained solving the following system of linear equations [54, 56]:

$$\begin{cases} \boldsymbol{\pi}\,\mathbf{\Lambda} = \mathbf{0} \\ \displaystyle\sum_{n=0}^{N-1} \pi_n = 1 \end{cases} \tag{4.16}$$

where the last equation specifies the normalization of the probability distribution.

Setting appropriately the matrix $\mathbf{\Lambda}$ with values extracted from a real breathing patient, it is possible to completely specify the CTMC model and employ it to replicate a RR pattern with the proper statistical behavior. Given that each patient can generate different RR patterns, depending on many factors, the infinitesimal generator matrix must be estimated. To this purpose, the estimation of $\mathbf{\Lambda}$ is carried out in three steps, as detailed in the following subsections: first, the RR pattern of the patient versus time is estimated;

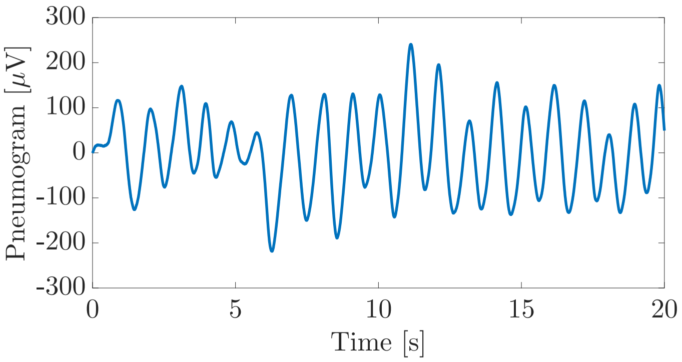


**Figure 4.5:** An example of a recorded pneumogram signal of a newborn patient kept under clinical observation.

second, the obtained pattern is fitted to a model with $N$ states and, finally, the transition rates are estimated from the finite-state pattern obtained at the previous step.

### 4.3.2 Estimation of Respiratory Rates

The RR is estimated by processing the pneumogram signal, which records the volume changes of the thoracic cavity of a patient and is obtained by placing an elastic belt around the chest. In Figure 4.5, an illustrative example of the pneumogram signal relative to a newborn patient breathing with RR of approximately 60 bpm is shown. Since the pneumogram describes the whole movements related to breathing, an algorithm to estimate the RR from this signal is needed. Excluding possible respiratory pauses or macromovements of the patient under observation, the pneumogram signal is a quasi-periodic signal. A method to estimate the RR from the pneumogram signal thus relies on the estimation of its fundamental frequency.

To this purpose, the pneumogram signal can be modeled by the following discrete-time signal:

$$p\left[i\right] = c + A\cos\left[2\pi\frac{f_0}{f_s}i + \phi\right] + w\left[i\right] \tag{4.17}$$

where $c$ is a constant component, $A$ is the amplitude of the periodic component, $f_0$ is the fundamental frequency, $f_s$ is the sampling frequency, $\phi$ is the phase and $\{w[i]\}$ is a sequence of i.i.d. zero-mean Gaussian noise samples. The main goal is to estimate the fundamental frequency $f_0$, which can be interpreted as the RR. This model is similar to that employed in Chapter 3.

To estimate $f_0$, the Maximum Likelihood (ML) criterion on a window of signal samples can be applied. Defining the parameter vector $\boldsymbol{\theta} = [A, f_0, \phi]$, the likelihood function $J(\boldsymbol{\theta})$ to be minimized is (see, e.g., Subsection 3.2.1 or [46]):

$$J(\boldsymbol{\theta}) = \sum_{i=0}^{M-1} \left( p[i] - A \cos\left[ 2\pi \frac{f_0}{f_s} i + \phi \right] \right)^2 \tag{4.18}$$

where $\{p[i]\}_{i=0}^{M-1}$ are now the observed samples in the considered $M$-sample time window. Following the steps introduced in (3.6)–(3.8), the approximate ML fundamental frequency estimator is based on the Discrete Fourier Transform (DFT) of the observed $M$ samples. Once $\hat{f}_0$ has been obtained, it is possible to estimate $A$ as in (3.8). If the estimated RR is below the value $R_L$ or the amplitude $\hat{A}$ is lower than a suitable threshold, absence of breathing is assumed and $\hat{f}_0$ is set to 0. The frequency $\hat{f}_0$ is finally declared as the RR of the patient in the observed window.

Since the pneumogram signal can be significantly noisy because of possible patient's movements or artifacts involving other body parts, which are associated with state $S_{N-1}$, as described at the beginning of Subsection 4.3.1, a preliminary time-domain check on the observed samples $\{p[i]\}_{i=0}^{M-1}$ is performed to detect such conditions. To this purpose, the analyzed window of samples of the pneumogram signal is checked against the condition

$$\exists i : \ |p[i]| > \eta \qquad i = 0, 1, \ldots, M-1 \tag{4.19}$$

where $\eta$ is a threshold set to distinguish respiratory movements from other ones. Then the estimated frequency $\hat{f}_0 = R_M$ is formally assigned so that $\varrho_{N-1} = R_M$.

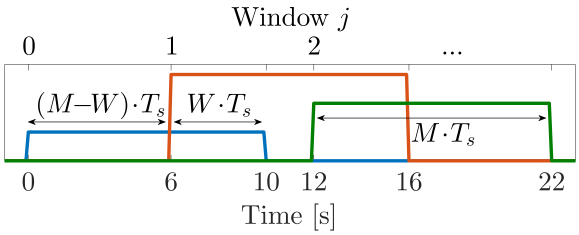


**Figure 4.6:** An illustrative example of interlaced windows, with a window length of $MT_s = 10$ s and interlacing interval $WT_s = 4$ s. Window analysis is carried out every $(M - W)T_s = 6$ s.

In order to obtain a RR pattern which represents the fundamental frequency $\hat{f}_0$ over time, the ML estimation is repeated over successive windows. Interlaced observation windows, with an interlacing interval of $W$ samples, allow to carry out the estimation every $M - W$ samples. Figure 4.6 shows an example of interlacing, with $(W/M) \cdot 100 = 40\%$ overlap and three consecutive windows. The integer $j$ specifies the window index. Estimating the RR along the pneumogram by the approach described above, it is possible to obtain a discrete-time signal $\hat{X}_c\,[j]$, representing the time evolution of the RR, where $\hat{X}_c\,[j]$ is defined as the fundamental frequency estimated in the $j$-th analyzed window by the above procedure. Therefore, the RR is estimated every $(M - W)\,T_s$ seconds, where $T_s = 1/f_s$ is the sampling interval of the pneumogram.

In order to quantize the estimated continuous-value RR into a finite state space, so that the statistical model introduced in Subsection 4.3.1 can be used, a proper quantization of the continuous-value discrete-time signal $\hat{X}_c\,[j]$ is needed as discussed in the following subsection.

### 4.3.3 Selection of Model States

The first step to fit the estimated signal $\hat{X}_c\,[j]$ into the model previously described is to define the $N$ states in $\mathcal{S}$ by selecting appropriate RR values

$\{\varrho_n\}_{n=0}^{N-1}$. An automatic method to select these values is used, with specific features depending on the presence of apnea events or random large body movements which affect the pneumogram signal. Following the different modeling cases described at the beginning of Subsection 4.3.1, various signal quantization methods are considered.

- If apnea events are not of interest and there are no patient random movements, the $N$ RR values of the model can be selected by the use of the Lloyd-Max algorithm [58, 59], which minimizes the mean square distortion between the signal $\hat{X}_c[j]$ and the $N$-state quantized one in the range of interest $[R_L, R_H]$.

- If apnea events are of interest, the method is modified as follows: a first state $S_0$ with rate $\varrho_0 = 0$ is assigned to describe absence of breathing; the remaining $N-1$ states $\{S_1, S_2, \ldots, S_{N-1}\}$ are estimated by the same Lloyd-Max algorithm used above. In this case, the overall range of interest becomes $\{0\} \cup [R_L, R_H]$.

- If random body movements are considered, the algorithm assigns the $N$-th state $S_{N-1}$ with $\varrho_{N-1} = R_M$ to time intervals in which the patient is moving; the remaining $N-1$ states $\{S_0, S_1, \ldots, S_{N-2}\}$ are estimated by the Lloyd-Max algorithm. The overall range becomes $[R_L, R_H] \cup \{R_M\}$.

- If both apneas/respiratory pauses and large random movements are considered, the algorithm assigns the state $S_0$ with rate $\varrho_0 = 0$ or the state $S_{N-1}$ with $\varrho_{N-1} = R_M$ to time intervals in which absence of breathing is detected or the patient is moving, respectively; the remaining $N-2$ states $\{S_1, S_2, \ldots, S_{N-2}\}$ are estimated by the Lloyd-Max algorithm. The overall range becomes $\{0\} \cup [R_L, R_H] \cup \{R_M\}$.

Once the RRs $\{\varrho_n\}$ and the corresponding states in $\mathcal{S}$ are defined, the pneumogram signal is quantized to the nearest RR value $\varrho_n$ present in the

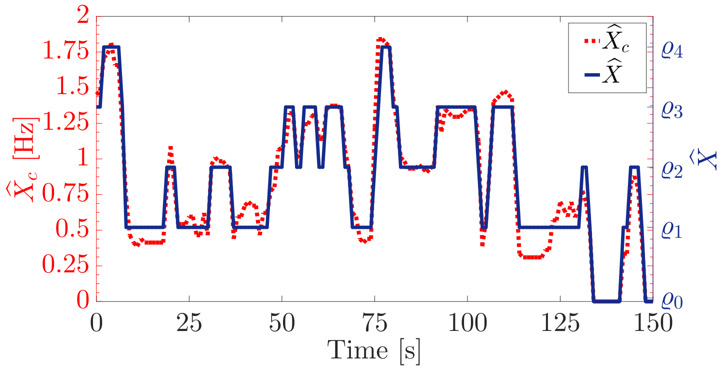


**Figure 4.7:** An illustrative example of the RR signal $\hat{X}_c$ estimated from the pneumogram and a quantized signal model $\hat{X}$ with $N = 5$ states; in this examples, a patient with possible respiratory pauses is considered.

model, thus obtaining the following discrete-value version of the signal $\hat{X}_c\,[j]$:

$$\hat{X}\,[j] \quad = \underset{\varrho_n \in \{\varrho_0, \varrho_1, \dots, \varrho_{N-1}\}}{\operatorname{argmin}} \left| \hat{X}_c\,[j] - \varrho_n \right| . \tag{4.20}$$

In Figure 4.7, an illustrative example of the RR pattern $\hat{X}_c[j]$, estimated by the method described in Subsection 4.3.2, together with the corresponding quantization $\hat{X}[j]$, computed according to (4.20), are shown. In this example, the estimation of the RR is carried out every second, on a temporal window of $MT_s = 10$ s and with window overlapping equal to 90% (i.e., $WT_s = 9$ s).

### 4.3.4 Infinitesimal Generators Estimation

Given the $N$-state model and the quantized RR pattern extracted from a sample patient, the description of the statistical model of the RR requires the definition of the transition rate matrix $\mathbf{\Lambda}$. The estimation of the infinitesimal generator matrix of CTMCs is a known problem, which becomes more difficult if the estimation is carried out on incomplete data or on sampled time series [60]. To simplify the discussion, $\hat{X}[j]$, whence the matrix $\mathbf{\Lambda}$ is estimated, is approximated as a continuous-time equivalent signal $\hat{X}(t)$—this assumption is valid, provided the RR is estimated with sufficiently high frequency, i.e., with a small $(M - W)\, T_s$ factor (e.g., $(M - W)\, T_s \leq 1$ s).

Owing to this simplification, the estimation of the transition rates is based on the ML estimation method [60, 61]. Assuming the process $\hat{X}(t)$ is observed in the interval $[0, T]$, the time spent by the process in state $S_n$ can be denoted by the random variable $R_n\,(T)$; similarly, the number of transitions from state $S_m$ to state $S_n$ in this observation interval can be denoted as $N_{m,n}\,(T)$. The log-likelihood function to derive an estimate $\hat{\mathbf{\Lambda}}$ of the infinitesimal generator matrix of the observed process $\hat{X}(t)$ is given by [61]

$$\log\left[\mathcal{J}\left(\mathbf{\Lambda}\right)\right] = \sum_{m=0}^{N-1} \sum_{\substack{n=0 \\ n \neq m}}^{N-1} \Big[ N_{m,n}\,(T) \log\left(\lambda_{m,n}\right) - \lambda_{m,n} R_n\,(T) \Big]. \tag{4.21}$$

By straightforward manipulations [60], the estimate of $\lambda_{m,n}$ can be expressed as

$$\hat{\lambda}_{m,n} = \frac{N_{m,n}\,(T)}{R_n\,(T)} \qquad \text{for } m \neq n. \tag{4.22}$$

The remaining rates for $m = n$ are obtained using (4.13).

Provided the observation interval $[0, T]$ of the RR pattern is sufficiently long, the ML approach allows a reliable estimation of the infinitesimal generator matrix $\mathbf{\Lambda}$.

### 4.3.5 Relationship with the Model for Apneas

The statistical model described in this section can be used to simulate apnea episodes and respiratory pauses, provided the state $S_0$ is associated with the rate $\varrho_0 = 0$. A simple statistical model, equal to the one in Section 4.2 with two-state CTMC ($N = 2$), can be employed where states $\{S_0, S_1\}$ describe the presence of an apnea/respiratory pause and regular breathing, respectively. Corresponding sojourn times $\{\tau_\ell\}$ denote the durations of respiratory pauses and normal breathing, conditionally on the corresponding state. The estimation method of the $2 \times 2$ infinitesimal generator matrix is used with only two states: $\bar{\varrho}_n = 1$ if regular breathing occurs and $\bar{\varrho}_n = 0$ in the presence of apneas/respiratory pauses.

## 4.4 Validation of Statistical Models

The validation of the two-state CTMC model is first discussed. Vital sign recordings were provided by the Neonatal Intensive Care Unit (NICU) of the University Hospital of Parma.

By using the model described Section 4.2, a 1-hour simulation of a newborn suffering from apnea events has been obtained, with features extracted from real data, of the same length, of an infant with CCHS. As ground truth, the results obtained by monitoring the infant by video-ElectroEncephaloGram (EEG), lasting a total of 1 h and 35 s, in which a total of 39 respiratory pauses are detected, are considered. These 39 episodes have a total duration of 31 min and 30 s; the average duration $\bar{a}$ of apnea episodes is 48 s and the average duration of regular breathing periods $\bar{b}$ between consecutive apnea episodes is 43 s. Despite these values may appear too large to be realistic, they were estimated based on the video-EEG, an EEG device which simultaneously record a video of the patient, whereas only 7 of the detected apneas were accompanied by severe desaturation and oxygen therapy was applied using nasal cannulas. As a comparison, Table 4.1 shows the values of mean sojourn times $\bar{a}$ and $\bar{b}$ obtained from 10 different simulations according to the proposed two-state model.

In Figure 4.8, histograms of the empirical distributions of (a) $a_i$ and (b) $b_i$, obtained by simulating apneas in a test lasting 24 h, are shown. Parameters $\mu$ and $\nu$ are properly set so that mean sojourn times correspond to those provided by the ground truth data mentioned above ($1/\mu = 43$ s and $1/\nu = 48$ s). The accuracy of the proposed exponential approximation, i.e., the validity of the proposed CTMC model, is evident.

The multiple-state CTMC is now validated. As examples, two recorded pneumogram samples of two different newborns, the first suffering from apnea

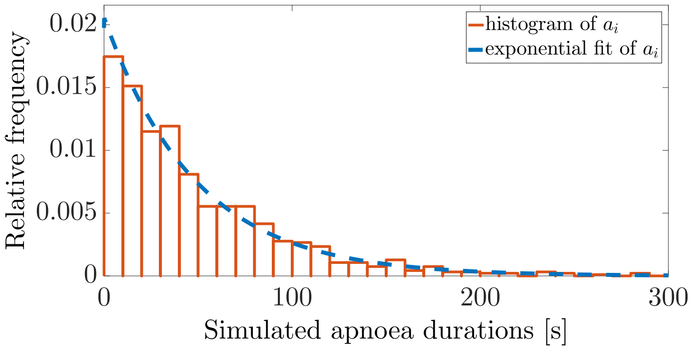
| **Simulation Number** | $\bar{a}$ [s] | $\bar{b}$ [s] | **Simulation Number** | $\bar{a}$ [s] | $\bar{b}$ [s] |
|---|---|---|---|---|---|
| 1 | 45.9 | 47.4 | 6 | 39.9 | 52.9 |
| 2 | 46.8 | 54.4 | 7 | 49.7 | 34.2 |
| 3 | 46.4 | 40.4 | 8 | 43.1 | 43.0 |
| 4 | 49.3 | 28.1 | 9 | 48.8 | 38.2 |
| 5 | 47.2 | 34.8 | 10 | 45.4 | 43.4 |

**Table 4.1:** Tests of simulated average sojourn times for the two-state CTMC.

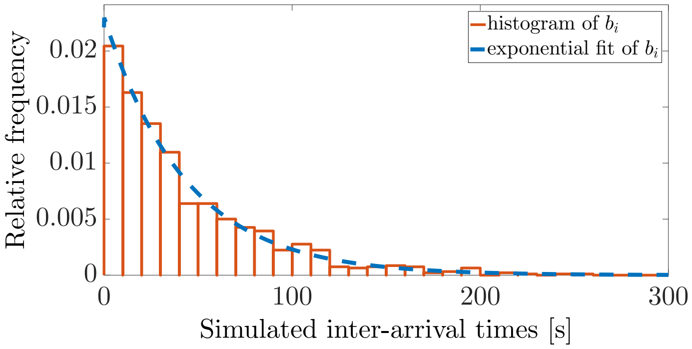
events and the second regularly breathing, are used for the RR data extraction: the first record has a total length of 1 h and 42 min and the second one of 1 h and 6 min. The fundamental frequency is estimated on temporal windows of duration $MT_s = 10$ s (with $T_s = 31.25$ ms) and with a 95% overlapping, namely $WT_s = 9.5$ s. Moreover, a heuristic threshold $\eta = 400$ $\mu$V is set in (4.19) to detect random movements to be assigned to the state $S_{N-1}$—this value of $\eta$ has been experimentally optimized. The threshold value is selected by observation of pneumogram signals of several newborn patients.

Fixing, as an example, the desired number of states of the model to $N = 5$, the automatic state selector, described in Subsection 4.3.3, extracts the RR sets $\{0, 0.5, 0.9, 1.3, R_M\}$ and $\{0.44, 0.74, 1.04, 1.33, R_M\}$ for the first and the second considered patients, respectively. Afterward, relying on the methods described in Subsection 4.3.4, the algorithm derives an estimate $\hat{\mathbf{\Lambda}}$ of the infinitesimal generator matrix. In Tables 4.2(a) and 4.2(b) the RR sets $\{\varrho_n\}$, the ML-estimated matrices $\hat{\mathbf{\Lambda}}$ and the corresponding stationary distributions $\boldsymbol{\pi}_{\mathrm{ML}}$ derived according to (4.16) for the two considered examples, are reported. It can be noticed that both patients are affected by random body movements which cause the presence of the state with $\varrho_{N-1} = R_M$.

It must be remarked that the description of large body movements or possible artifacts with a specific state with rate $\varrho_{N-1} = R_M$ in the CTMC is fundamental in order to avoid degradation of the statistical behavior of the



(a)



(b)

**Figure 4.8:** Empirical probability distributions of (a) $a_i$ and (b) $b_i$, with corresponding estimated average values of $\bar{a} = 48.6$ s and $\bar{b} = 43.4$ s, respectively.

model. In fact, if this state was not available, faulty estimation of the RRs might arise, with possible wrong selection of RR values $\{\varrho_n\}$ and incorrect estimates of transition rates and infinitesimal generator matrix.

In Figure 4.9, comparisons between the histograms of estimated frequencies $\hat{X}[j]$, whose Probability Mass Function (PMF) is defined as the vector $\mathbf{p}_{\hat{X}}$, and the stationary distributions $\boldsymbol{\pi}_{\text{ML}}$ relative to the ML-estimated matrix $\hat{\boldsymbol{\Lambda}}$ for both examples are shown: Part (a) corresponds to the first example of the patient suffering from apneas and Part (b) is related to the second example of

| **Example 1 (patient suffering from apnea events)** |
|---|
| $\{\varrho_n\} = \{0 \quad 0.5 \quad 0.9 \quad 1.32 \quad R_M\}$ |
| $\hat{\mathbf{\Lambda}} = \begin{bmatrix} -0.188785 & 0.08972 & 0.04486 & 0.042991 & 0.011215 \\ 0.047083 & -0.203685 & 0.073695 & 0.071648 & 0.011259 \\ 0.014614 & 0.080376 & -0.22547 & 0.127349 & 0.003132 \\ 0.007717 & 0.019756 & 0.036734 & -0.066677 & 0.00247 \\ 0.042272 & 0.02642 & 0.002642 & 0.002642 & -0.073976 \end{bmatrix}$ |
| $\boldsymbol{\pi}_{\mathrm{ML}} = [0.08788 \quad 0.16048 \quad 0.15736 \quad 0.53211 \quad 0.06217]$ |

(a)

| **Example 2 (patient regularly breathing)** |
|---|
| $\{\varrho_n\} = \{0.44 \quad 0.74 \quad 1.04 \quad 1.33 \quad R_M\}$ |
| $\hat{\mathbf{\Lambda}} = \begin{bmatrix} -0.205567 & 0.124197 & 0.059957 & 0.004283 & 0.017131 \\ 0.020036 & -0.080859 & 0.047943 & 0.003578 & 0.009302 \\ 0.014957 & 0.071581 & -0.108974 & 0.014957 & 0.007479 \\ 0.015873 & 0.047619 & 0.206349 & -0.317460 & 0.047619 \\ 0.003656 & 0.007313 & 0.006399 & 0 & -0.017367 \end{bmatrix}$ |
| $\boldsymbol{\pi}_{\mathrm{ML}} = [0.05644 \quad 0.32998 \quad 0.22677 \quad 0.01516 \quad 0.37164]$ |

(b)

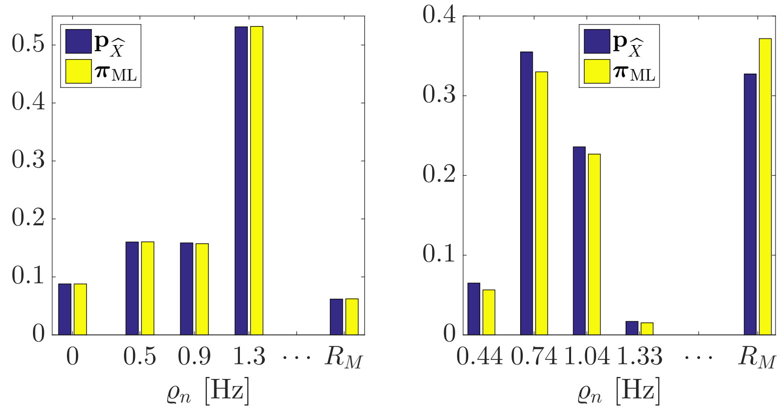
**Table 4.2:** Estimated RR sets, infinitesimal generator matrices and corresponding stationary distributions of (a) a newborn patient suffering from apnea events and (b) a newborn regularly breathing.

the regularly breathing newborn.

In order to quantitatively compare the similarity of the two PMFs in Figure 4.9, the Kullback-Leibler (KL) divergence [62] may be used:

$$D_{\mathrm{KL}}\left(\mathbf{p}\|\mathbf{q}\right) = \sum_n p_n \log_2 \frac{p_n}{q_n} \tag{4.23}$$

in which $p_n$ and $q_n$ denote the probability masses of the distributions $\mathbf{p}$ and $\mathbf{q}$, respectively. This quantity is a measure of the difference between the “true” distribution $\mathbf{p}$ and the “assumed” distribution $\mathbf{q}$; it is expressed in bits, due to the use of $\log_2$. In both examples shown in Figure 4.9, the KL divergence



(a) (b)

**Figure 4.9:** Comparisons between histograms with automatically selected bins for the process $\hat{X}$ (pneumogram-based) with PMF $\mathbf{p}_{\hat{X}}$ and the stationary distributions $\boldsymbol{\pi}_{\mathrm{ML}}$ relative to (a) the example of the newborn patient suffering from apnea events and (b) the regularly breathing newborn.

$D_{\mathrm{KL}}\left(\mathbf{p}_{\hat{X}} \| \boldsymbol{\pi}_{\mathrm{ML}}\right)$ is computed: values of $0.011 \cdot 10^{-3}$ bits and of $6.642 \cdot 10^{-3}$ bits are obtained for the cases (a) and (b), respectively. The stationary distributions are very similar to the histogram-based ones, with very low KL divergence values in both examples, confirming that the CTMC model has a steady state behavior similar to that of the RR pattern of the real patients.

To further verify the effectiveness of the statistical model, the effects of varying the number of states of the CTMC is now discussed. For this validation example, a new patient is considered: the total length of the recorded pneumogram signal is 51 min. In Tables 4.3(a), 4.3(b) and 4.3(c) the extracted RR sets $\{\varrho_n\}$, corresponding estimated infinitesimal generator matrices and stationary distributions $\boldsymbol{\pi}_{\mathrm{ML}}$ with $N = 4$, 5 and 6 are reported, respectively. Changes in the transition rates when the number of states of the model is varied can be appreciated—the use of a larger number of levels allows a finer representation of the RR values. Furthermore, as the number of the states $N$ increases, and the set $\{\varrho_n\}$ changes, the infinitesimal generators and the sta-

| **CTMC model with $N = 4$** |
|---|
| $\{\varrho_n\} = \{0.49 \quad 0.88 \quad 1.27 \quad R_M\}$ |
| $\hat{\boldsymbol{\Lambda}} = \begin{bmatrix} -0.196532 & 0.109827 & 0.023121 & 0.063584 \\ 0.021876 & -0.058898 & 0.012621 & 0.024401 \\ 0 & 0.105263 & -0.144044 & 0.038781 \\ 0.006814 & 0.021124 & 0.003407 & -0.031346 \end{bmatrix}$ |
| $\boldsymbol{\pi}_{\mathrm{ML}} = \{0.0605 \quad 0.38942 \quad 0.05555 \quad 0.49453\}$ |

(a)

| **CTMC model with $N = 5$** |
|---|
| $\{\varrho_n\} = \{0.44 \quad 0.73 \quad 1.03 \quad 1.32 \quad R_M\}$ |
| $\hat{\boldsymbol{\Lambda}} = \begin{bmatrix} -0.227545 & 0.083832 & 0.023952 & 0.023952 & 0.095808 \\ 0.011161 & -0.071429 & 0.035714 & 0.001116 & 0.023438 \\ 0.010604 & 0.067869 & -0.123012 & 0.012725 & 0.031813 \\ 0 & 0.021978 & 0.076923 & -0.131868 & 0.032967 \\ 0.003406 & 0.014986 & 0.011580 & 0.001362 & -0.031335 \end{bmatrix}$ |
| $\boldsymbol{\pi}_{\mathrm{ML}} = \{0.02902 \quad 0.29355 \quad 0.15485 \quad 0.02781 \quad 0.49477\}$ |

(b)

| **CTMC model with $N = 6$** |
|---|
| $\{\varrho_n\} = \{0.41 \quad 0.65 \quad 0.88 \quad 1.11 \quad 1.35 \quad R_M\}$ |
| $\hat{\boldsymbol{\Lambda}} = \begin{bmatrix} -0.263566 & 0.093023 & 0.015504 & 0.015504 & 0.031008 & 0.108527 \\ 0.015102 & -0.133765 & 0.084142 & 0.006472 & 0.002157 & 0.025890 \\ 0.005735 & 0.061649 & -0.131900 & 0.035842 & 0 & 0.028674 \\ 0.007828 & 0.003914 & 0.101761 & -0.164384 & 0.027397 & 0.023483 \\ 0 & 0 & 0.067227 & 0.100840 & -0.201681 & 0.033613 \\ 0.002723 & 0.008850 & 0.013615 & 0.005446 & 0.000681 & -0.031314 \end{bmatrix}$ |
| $\boldsymbol{\pi}_{\mathrm{ML}} = \{0.02143 \quad 0.1547 \quad 0.22708 \quad 0.08513 \quad 0.01818 \quad 0.49348\}$ |

(c)

**Table 4.3:** Estimated RR sets and infinitesimal generator matrices for a newborn patient with different values of the number of states $N$; in particular (a) with $N = 4$ states, (b) with $N = 5$ and (c) with $N = 6$ states.

tionary distribution vary and are "redistributed" on the new states; for what concern the state $R_M$, it can be noticed that the values of $\lambda_{N-1,N-1}$ remain approximately the same, as it occurs to the values of $\pi_{N-1}$.

## 4.5 Conclusion

In this chapter, possible solutions to the problem of developing statistical models which describe breathing patterns, respiratory pauses and apneas, are proposed.

First, a simple CTMC-based model to replicate the behavior of respiratory pauses/apnea events was described. It relies on a two-state CTMC, where the two states represent normal breathing and absence of respiration, respectively. This model was then extended to a multiple-state model that is able to represent complete patterns of the RRs of a patient.

The problem of selecting the parameters of statistical models is then tackled. To this purpose, an approach to estimate the parameters characterizing the CTMC from real medical data for each one of the developed models is proposed.

Finally, both statistical models are tested and validated by simulating various kinds of patients, regularly breathing and affected by apnea events, commenting the obtained results and showing the effectiveness of the developed simulators.

# Chapter 5

# Simulators of Apneas and Breathing Patterns

## 5.1 Introduction

In Chapter 4, theoretical statistical models of breathing patterns and specific breathing disorders were introduced. The two models previously described can be employed to develop realistic simulators, which can be useful to create large databases of video recordings which allow to develop, test and optimize video processing-based algorithms for remote monitoring of respiration and related diseases.

In the literature, some solutions for the simulation of the behavior of anatomical or biological functions are presented. Some physically-based anatomical simulators are first introduced. In [63], a hardware system to handle bio-mechanical movements and simulate an anatomical and functional model of the evolution of the human trunk structures during respiration is proposed; in [64], a system of rigid and deformable parts, which simulates the biological function of respiration for computer animation, is presented. Then, software simulators are also considered. In [65], a software-based audio simulation model of snoring signals is proposed to provide a repeatable test-bed for signal processing techniques to aid clinical diagnosis of breathing abnormalities during sleep. Video simulation of motion is also used in different contexts, e.g. to simulate dynamics in a still flame [66] or to simulate subtle motions, due to natural forces, on passive elements in still pictures [67].

In this chapter, the statistical models described in the previous Chapter 4 are employed as background for the definition of two types of simulators. *Software-based* simulators, able to directly manipulate video recordings of regularly breathing patients in order to introduce artificial respiratory disorders and modify the breathing pattern, are first presented. A simple *hardware-based* simulator is also developed: it exploits a manikin equipped with a moving chest to physically reproduce possible breathing disorders according to the proposed statistical model. Developed simulators are then used to test extensively video processing-based algorithms for Respiratory Rate (RR) monitoring and apnea detection.

This chapter is organized as follows. In Sections 5.2, two software-based simulators are presented. The first one, described in Subsection 5.2.1, is able to insert respiratory pauses/apneas in video streams framing a regularly breathing patient, without modifying his normal breathing pattern. The second one, described in Subsection 5.2.2, can create a simulated breathing pattern by processing a video recording of a patient with constant RR. Then, in Section 5.3, a physically-based simulator, reproducing thoracic movements of a patient is described. The assessment of the performance of video processing-based algorithms introduced in the previous Chapter 3, specifically the ones described in Sections 3.2 and 3.3, are addressed in Section 5.4. Finally, Section 5.5 is dedicated to final remarks.

## 5.2 Video Processing-based Simulators

In Chapter 4, a simple model for apnea/respiratory pauses and an extended one for complete breathing patterns, were presented. Following this order, two software-based simulators, able to properly modify video recordings, are here introduced. A first one keeps unchanged the breathing pattern of the framed patient and inserts artificial apnea events inside the video recordings. The second one, starting from a video of a regularly and constantly breathing patient, can create a realistic breathing pattern, varying the RR of the framed patient.

### 5.2.1 Apnea Simulator

A video stream with $L$ frames is assumed as data source. The apnea simulator has two main tasks:

- once the parameters $\{\mu, \nu\}$ have been estimated, it generates durations of normal breathing periods $\{b_i\}$ and the durations of respiratory pauses/apneas $\{a_i\}$ following the two-state Continuous-Time Markov Chain (CTMC) model defined in Subsection 4.2;
- it derives a new video stream with incorporated apneas/respiratory pauses specified by the times $\{b_i\}$ and the durations $\{a_i\}$, generated in the previous step.

Thus, the first step consists in extracting the sojourn times for the two states of the model. Assuming that the initial state of the patient at time $t = 0$ is unknown, the state is chosen according to a Bernoulli random variable [54] with probabilities specified by the steady-state distribution of the CTMC in (4.8). Once the initial state is set, the first sojourn time is generated according to the proper exponential distribution in (4.4). If the initial state is $S_1$, $b_1$ is generated; otherwise, $a_1$ is generated and $b_1$ is set to 0. Knowing the video frame rate $f_r$ of the video input, the continuous sojourn times can be expressed in terms of numbers of frames according to

$$\begin{aligned} \tilde{b}_i &= \text{round}\left(b_i f_r\right) \\ \tilde{a}_i &= \text{round}\left(a_i f_r\right) \end{aligned} \tag{5.1}$$

where the round$(\cdot)$ operator denotes the integer closest to the argument (rounding function). Therefore, the "discretized" CTMC waits $\tilde{b}_i$ or $\tilde{a}_i$ frames

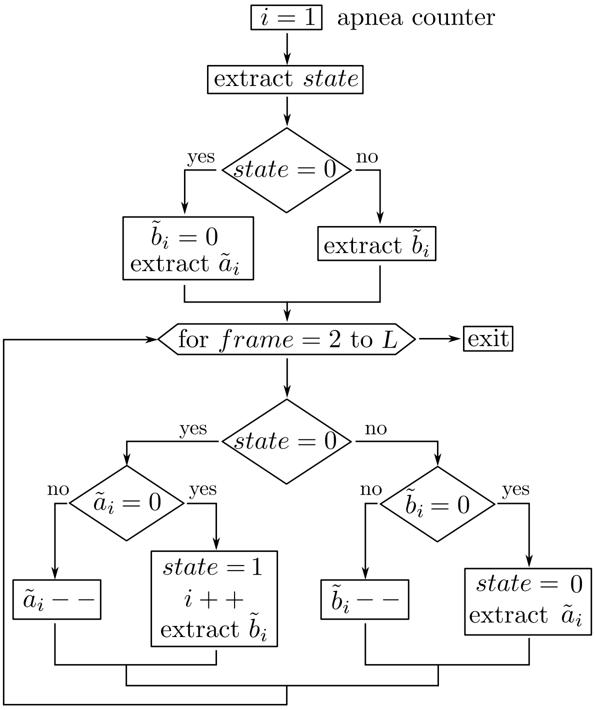


**Figure 5.1:** Flow diagram of the algorithm used to implement the two-state CTMC model of the respiratory process.

before switching to the other state: when a state change occurs, the algorithm generates the sojourn time of the new state. This process is repeated until the end of the $L$ frames of the input video stream. In Figure 5.1, a flow diagram of the algorithm used to implement the CTMC model, is shown: the upper part describes the selection of the initial state, whereas the lower one refers to the cycles for the selection of durations of normal breathing periods and apneas. The outputs of the two-state CTMC simulator consist of the actual sequences $\{b_i\}$ and $\{a_i\}$ which will be used by the apnea inserter.

After the generation of sojourn times by means of the two-state CTMC model, the simulator begins to produce a new video where the artificially generated apneas are inserted. The system scans the whole video sequence inserting apneas with frame durations $\{\tilde{a}_i\}$, spaced by $\{\tilde{b}_i\}$. An apnea of duration $\tilde{a}_i$ is generated extracting, first, a block of $\tilde{\Delta}_i < \tilde{a}_i$ frames from the original video, starting from the $\tilde{t}_i$-th frame, where

$$\begin{aligned} \tilde{t}_i &= \mathrm{round}\,(t_i f_r) \\ \tilde{\Delta}_i &= \left\lceil \frac{a_i}{c} f_r \right\rceil \end{aligned} \tag{5.2}$$

in which $\lceil \cdot \rceil$ denotes the smallest integer larger than or equal to the argument (ceiling function) and the parameter $c > 1$ is heuristically set. In order to insert an apnea, the simulator makes use of a pixel-wise interpolation of the extracted video block in order to expand the $\tilde{\Delta}_i$ frames to the corresponding sojourn time $\tilde{a}_i$. In other words, the video block is "stretched", by the factor $c$, to a proper apnea length by a pixel-by-pixel processing with a shape-preserving piecewise cubic interpolation—specifically, using a piecewise cubic hermite interpolating polynomial algorithm [68]. Interpolation is used with the purpose to slow down breathing movements and to simulate an apnea event in the monitored patient.

The noise caused by camera sensors and wire connections present in the original video is also subject to interpolation: this means that the noise slows down together with the breathing movements. In order to preserve noise characteristics similar to the original ones, a noise compensating algorithm has been devised. Assuming that this noise is uniformly distributed inside the frames, it is possible to estimate its characteristics by processing the pixels with static background. In particular, choosing border pixels of the input video and assuming static background (static camera), the system estimates the variance for every border pixel with an unbiased variance estimator [46]. Upon estimation of the variance of the background pixels, a variance value is derived and used to generate the noise to be inserted in the video stream. In order to update the camera noise in the new interpolated video block, a white Gaussian noise sequence, with zero mean and variance set to the one previously estimated, is added to all pixels.

As a result, the interpolated video (of duration $\tilde{a}_i$ frames) is inserted into the output video stream starting from $\tilde{t}_i$. At frame $\tilde{t}_i + \tilde{\Delta}_i$, the system switches to state $S_0$ and waits for the next apnea event. The overall procedure is repeated until all apneas to be simulated have been generated. The global frame length of the output video sequence with simulated apneas can thus be expressed as follows:

$$L' = L + \sum_j \left( \tilde{a}_j - \tilde{\Delta}_j \right). \tag{5.3}$$

### 5.2.2 Simulator of Breathing Patterns

The simulator presented in the previous subsection is only able to insert respiratory pauses/apneas in video streams, but not to create a custom breathing pattern. The here proposed video-based simulator is able to process a video of a regularly breathing patient and modify it by creating a new video with variable RRs relying on the statistical model introduced in Section 4.3. The RR of the framed patient in the source video is approximated as time-invariant and defined as $\varrho_V$. The rates associated with the various states can be described by the ratios between the rates of the corresponding state and the RR of the patient in the original video. The normalized RRs used in the model can be expressed as:

$$\bar{\varrho}_n = \frac{\varrho_n}{\varrho_V} \qquad n = 0, 1, \ldots, N-1. \tag{5.4}$$

After the waiting and jump times are generated according to the $N$-state CTMC, the corresponding number of frames are obtained similarly to (5.1):

$$\begin{aligned} \tilde{\tau}_\ell &= \mathrm{round}\left(\tau_\ell f_r\right) \\ \tilde{t}_\ell &= \mathrm{round}\left(t_\ell f_r\right) \end{aligned} \tag{5.5}$$

where $f_r$ is the frame rate of the video input. If the RRs $\{\varrho_n\}$ are all different from the rate $\varrho_V$ of the original video, it may be convenient to scale them by a factor $C_V$ chosen so that $\varrho_{\max} C_V = \varrho_V$, where $\varrho_{\max}$ is the RR corresponding to the state with lowest rate, i.e. $\min_{n \in \{0,1,\ldots,N-1\}} \mu_n$, and, consequently, longest mean sojourn time.

The simulator then starts producing a new video where the synthetic RR pattern is inserted. The system scans the whole video inserting breathing times with durations $\{\tilde{\tau}_\ell\}$. The $\ell$-th breathing time is simulated by processing a video block of $\tilde{d}_\ell$ frames starting from the $\tilde{t}_\ell$-th video frame, where

$$\tilde{d}_\ell = \left\lceil \frac{\tau_\ell}{\bar{\varrho}_n} f_r \right\rceil \tag{5.6}$$

and $\lceil\cdot\rceil$ denotes the ceiling function. A RR is generated according to the state to be simulated.

- If $S_n$ is such that $\bar{\varrho}_n = 1$, no video processing is needed, since the desired RR is equal to that of the patient breathing in the original video.
- If $S_n$ is such that $0 < \bar{\varrho}_n < 1$, a block of $\tilde{d}_\ell < \tilde{\tau}_\ell$ frames is extracted, in order to slow down breathing movements of the recorded patient; the extracted video block is "stretched" to the proper length $\tilde{\tau}_\ell$ by the use of pixel-wise interpolation in the temporal dimension by a cubic spline [69].
- If $S_n$ is such that $\bar{\varrho}_n > 1$, a block of $\tilde{d}_\ell > \tilde{\tau}_\ell$ frames is extracted, in order to speed up breathing movements of the recorded patient, "contracting" the video block to the proper length $\tilde{\tau}_\ell$ by decimation in the temporal dimension using a cubic spline [69].
- If $S_0$ is such that $\bar{\varrho}_0 = 0$, the procedure used for $0 < \bar{\varrho}_n < 1$ cannot be used; in this case, the value of $\bar{\varrho}_0$ is replaced by a new value $\bar{\varrho}'_0$ properly chosen such that $0 < \varrho'_0 \ll R_L$, where $\varrho'_0 = \bar{\varrho}'_0 \varrho_V$, so that the simulated RR is a convenient value clinically considered an apnea; then, the procedure illustrated for $0 < \bar{\varrho}_n < 1$ is applied with a modified $\bar{\varrho}'_0$ value; the results described in the next section use a value $\varrho'_0 \simeq 0.1$ Hz.

For $\bar{\varrho}_n < 1$, the noise problem described at the end of Subsection 5.2.1, manifests itself in this simulator in a similar manner. As in the previous case, to compensate for the noise not included because of the interpolation filter, a sequence of uncorrelated Gaussian samples with zero mean and variance $\hat{\sigma}^2$ is generated and then filtered by a high-pass filter with 3 dB cut-off frequency equal to that of the interpolating filter. As before, the variance $\hat{\sigma}^2$ can be estimated by background pixels. To update the camera noise inside the new video

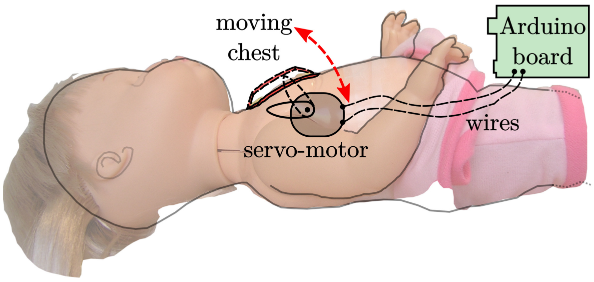


**Figure 5.2:** Illustrative diagram of the hardware-based simulator. The moving chest is driven by the servo-motor inside the manikin connected to a motor shield for an Arduino board.

block, the filtered noise sequence is added to all time-interpolated pixels. This procedure is not necessary for the cases with $\bar{\varrho}_n \geq 1$, because the decimation process does not modify the noise statistics. The overall procedure is repeated until the complete RR pattern has been inserted into the video.

## 5.3 Hardware-based Simulator

The hardware-based simulator consists of a manikin of an infant with moving parts able to mimic respiratory chest movements of the newborn. In Figure 5.2, an illustrative representation of the manikin simulator is shown. It consists of a moving chest coupled with the body of the manikin. The moving chest is driven by a mechanical arm connected to an electric servo-motor, inserted inside the body of the manikin, and controlled by a motor shield for Arduino UNO [70], a board based on an Atmel ATmega328P microcontroller. The controller is able to move the chest at a user-defined frequency with asymmetric speed to distinguish inhalation and exhalation movements. The servo-motor can vary the RR approximately from 2 bpm to 200 bpm, equivalent to a RR range between 0.033 Hz and 3.33 Hz, which readily allows to simulate the RRs of a newborn.

To simulate the respiratory behavior of a newborn, times $\{\tau_\ell\}$ and corresponding states with RRs $\{\varrho_n\}$ generated by CTMC-based simulation are passed to the microcontroller, which drives the servo-motor to mimic the RR pattern, moving the chest of the manikin with the selected RR for the required time. In the case of a respiratory pause or apnea event, the servo-motor is slowed down to the minimum reachable rate for a time duration equal to the sojourn time in the apnea state.

## 5.4 Performance of Video Monitoring Algorithms

After definitions of the statistical models in Chapter 4 and simulators in Sections 5.2 and 5.3, which employ CTMCs, these simulators are here validated and tested on previously developed video-based algorithms for the estimation of the RR and the detection of apnea events. Examples of application to assess the performance of systems for video-based RR estimation and apnea detection are presented.

The simulators have been used to generate videos of breathing patients with a statistically-defined RR pattern. The video-based simulator is used to process video streams recorded in the Neonatal Intensive Care Unit (NICU) of the University Hospital of Parma, with cameras having a frame rate $f_r = 25$ frame/s. Videos of the hardware-based simulator (i.e., the "breathing" manikin) are instead recorded with cameras operating at $f_r = 15$ frame/s. Both simulators are employed with the state set $\mathcal{S}$, RRs set $\{\varrho_n\}$ and matrix $\hat{\mathbf{\Lambda}}$ estimated in Table 4.2(a) and 4.2(b) for patients regularly breathing and affected by apnea events, respectively.

These simulators have been used to test and evaluate the performance of video processing-based systems to monitor the RR and detect apnea events. The obtained video sequences have been analyzed using algorithms developed in Chapter 3. In particular, the algorithm described in Section 3.2 and in [71] is referred to as Motion Magnification for Apnea Detection (MMAD). The algorithm described in Section 3.3 and in [72] is referred to as Spatio-Temporal video-processing for RR Estimation (STRE). Both MMAD and STRE algo-

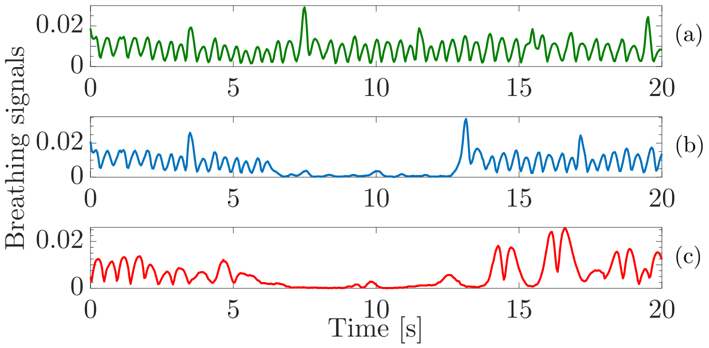


**Figure 5.3:** Examples of breathing signals of a newborn: (a) normal breathing pattern, (b) a software-simulated respiratory pause and (c) a real respiratory pause in the same patient, lasting approximately as the simulated one.

rithms can extract signals representative of breathing motion from a video stream—they are then used to detect apnea events or to estimate the RR. These algorithms analyze extracted breathing signals on temporal windows of 10 s with window interlacing equal to 90%.

In Figure 5.3, a direct comparison between motion signals extracted from the three possible videos using the STRE algorithm are depicted. In Part (a), the breathing signal extracted from a newborn without respiratory pauses is shown, where a RR approximately equal to 1.17 Hz can be recognized. In Part (b), the breathing signal, obtained by inserting a simulated respiratory pause into the video stream related to Part (a), is shown. In this example, the simulated pause begins at time instant 7.3 s and lasts approximately 6.2 s. For comparison purposes and to demonstrate the effectiveness of the software-based simulator, in Part (c) the breathing signal extracted from the video of a child suffering from Congenital Central Hypoventilation Syndrome (CCHS), containing a real respiratory pause lasting approximately 6.26 s, is shown. The similarity of the breathing signal in Figure 5.3(c) with the signal embedding the simulated respiratory pause in Figure 5.3(b) can be appreciated.

### 5.4.1 Detection of Apnea Events

Recall that apnea events are defined as episodes of complete absence of breathing lasting at least 20 s or between 10 s and 20 s, if they are associated with other clinic signs or symptoms; in the following, adopting a conservative approach, these two conditions are not distinguished and any episode of absence of breathing of at least 10 s is considered. The absence of breathing for less than 10 s is considered a respiratory pause and is not clinically relevant.

Results based on the simulator of Subsection 5.2.1 are first presented. The proposed simulator is employed to insert breathing pauses into a video, lasting 23 min and 45 s, showing a sleeping newborn. Two simulations were carried out, considering test values of $\mathbb{E}\{b_i\} = 60$ s ($\mu = 0.0167$ s$^{-1}$) and $\mathbb{E}\{a_i\} = 30$ s ($\nu = 0.0333$ s$^{-1}$). The obtained videos last, respectively, 30 min and 59 s and 30 min and 6 s. In the first video, the simulator added 19 respiratory pauses with a total duration of 7 min and 14 s, an average single respiratory pause duration of 21.7 s, a maximum duration of 101.2 s and a minimum duration of 2.6 s. In the second video, the simulator added 17 respiratory pauses with a total duration of 6 min and 21 s, an average single respiratory pause duration of 22.4 s, a maximum duration of 70 s and a minimum duration of 1.3 s.

These two videos, described in the previous paragraph and containing simulated respiratory pauses, are then processed by the automatic apnea detection system, previously described in Section 3.2, to test the quality of the simulated videos. The two simulated videos are obtained inserting randomly generated respiratory pauses into a single video of a normally breathing newborn. The automatic detection system focuses on the detection of respiratory pauses with duration of at least 10 s only. The obtained performance results are shown in terms of sensitivity ($100\alpha$) and specificity ($100\beta$) [73], as defined in (3.16) and expressed as percentage. Recall that the goal of the detection system is to identify apneas promptly, because potentially harmful. The sensitivity and specificity values obtained using the algorithm described in Section 3.2 for the first and the second video simulation are shown, respectively, in Tables 5.1 and 5.2, considering two different values of the Tolerance Delay (TD). In par-

| **Performance on Apnea Detection** | | |
|---|---|---|
| Tolerance Delay [s] | Sensitivity | Specificity |
| 0 | 77% | 90% |
| 10 | 89% | 98% |

**Table 5.1:** Video 1: Sensitivity and specificity performance

| **Performance on Apnea Detection** | | |
|---|---|---|
| Tolerance Delay [s] | Sensitivity | Specificity |
| 0 | 46% | 90% |
| 10 | 100% | 97% |

**Table 5.2:** Video 2: Sensitivity and specificity performance

ticular, $\alpha$ and $\beta$ are computed by considering the portions of minutes of apnea (out of 6 min and 50 s for the first video and 6 min and 21 s for the second one) and regular breathing (23 min and 45 s for both videos) correctly/incorrectly diagnosed. The algorithm allowed to detect: between 4 and 6 episodes, out of a total of 8, in the first video; and between 3 and 6 episodes, out of a total of 6, in the second video—the values depend on the acceptable TD in reporting an episode.

Finally, an extensive test of the simulators introduced in Section 5.2, the multi-state CTMC-based, specifically, and 5.3 is reported. The two simulators have been used to generate videos of newborns suffering from apnea events to test the previously proposed algorithms in Sections 3.2 and 3.3. The software-based simulator has been used to generate a video lasting approximately 1 h. The obtained video includes 74 simulated respiratory pauses, with 13 events, lasting at least 10 s each, which can be interpreted as apneas. The total duration of simulated apnea events is 166 s, with an average duration of 14.55 s and a maximum duration of 35 s. The hardware-based simulator is used to record a video lasting approximately 46 min. The simulation includes 33 simulated respiratory pauses, with 12 events lasting at least 10 s each. The total duration of simulated apnea events is 220 s, with an average duration of 17.08 s and a maximum duration of 33 s.

The obtained videos are processed by the algorithms MMAD and STRE, which implement automatic apnea detection systems. The two algorithms focus only on events lasting at least 10 s. The performance of these detection systems is investigated considering a binary test, which classifies results as "presence of apnea" (positive event) or "normal breathing" (negative event). The performance results are presented in terms of sensitivity and specificity, defined, respectively, as in (3.16). As a global measure of test performance, the Diagnostic Odds Ratio (DOR) [74] can also be employed, defined as

$$\Delta \triangleq \frac{T_{\mathrm{TP}}/T_{\mathrm{FN}}}{T_{\mathrm{FP}}/T_{\mathrm{TN}}} = \frac{\alpha}{1-\alpha} \cdot \frac{\beta}{1-\beta}. \tag{5.7}$$

In Figure 5.4, the performance results, in terms of Receiver Operating Characteristic (ROC) curves [75], are presented. The curves are obtained testing the algorithms for various values of the decision thresholds for presence/absence of periodicity related to the breathing signal. Specifically, in Figure 5.4(a) and Figure 5.4(b) the ROC curves for MMAD and STRE algorithms, tested on videos generated by software-based and hardware-based simulators, are shown, respectively. As concise performance indicator, in Figure 5.4 the values of the Area Under Curve (AUC) parameter [75] associated with the considered algorithms and simulators are also shown—the higher the AUC, the better the performance.

Optimum values of decision thresholds can be defined considering the point of the ROC curve with minimum Euclidean distance to the point $(0, 1)$ which describes the ideal detector. Considering optimum threshold values for both algorithms, sensitivity, specificity and DOR for the video streams obtained above are shown in Tables 5.3 and 5.4 for the software- and hardware-based simulators, respectively. These results show that the sensitivity is high for both algorithms—MMAD and STRE can effectively identify patients suffering from apneas. However, the specificity is higher with STRE than with MMAD—

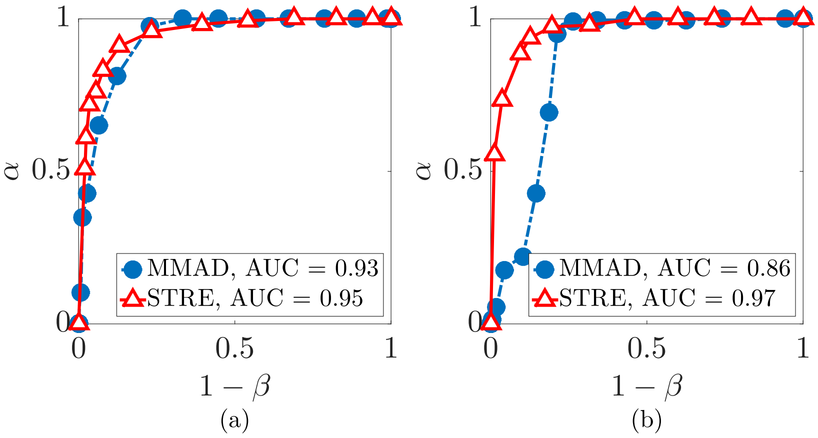


**Figure 5.4:** ROC curves for algorithms MMAD and STRE: (a) video-based simulator and (b) hardware-based simulator.

| **Algorithm** | Sensitivity | Specificity | $\Delta$ |
|---|---|---|---|
| MMAD | 88.8% | 82.9% | 38.4 |
| STRE | 91.0% | 86.9% | 67.1 |

**Table 5.3:** Detection performance on software-based simulator.

| **Algorithm** | Sensitivity | Specificity | $\Delta$ |
|---|---|---|---|
| MMAD | 95.1% | 78.7% | 71.7 |
| STRE | 92.3% | 89.6% | 103.3 |

**Table 5.4:** Detection performance on hardware-based simulator.

STRE can identify patients breathing normally more efficiently than MMAD. The global measure DOR has clearly higher values for STRE, indicating the better overall performance of this algorithm with respect to MMAD.

The presented results highlight the importance of the proposed breathing statistical models and simulators for performance analysis and optimized design of video-based monitoring systems.

### 5.4.2 Respiratory Rate Estimation

Video processing-based RR estimators can be tested comparing estimated rates with those simulated by the statistical model. The performance of the STRE algorithm in RR estimation is here assessed considering two performance metrics: (i) the Root Mean Squared Error (RMSE) between the simulated rate and that estimated by the video processing-based algorithm and (ii) the probability of correct estimation of RR, defined, according to medical practice, by the condition that the RR falls inside a tolerance range of ±15% with respect to the correct true value.

First, a video with a total length of 17 min and 54 s, generated by the software-based simulator, is analyzed. The simulated video is obtained by processing a video sample of a sleeping newborn breathing regularly with an approximate rate $\varrho_V = 0.69$ Hz. The RR is correctly estimated in 940 out of 1074 temporal windows (i.e., with a probability of correct estimation equal to 0.875), with a RMSE equal to 0.063 Hz. The RMSE, normalized with respect to the Root Mean Squared (RMS) value of the simulated RR, is about 10.2%.

Then, a video sample of the hardware-based simulator is analyzed: the recording has a total length of 8 min and 48 s, during which the manikin simulates a breathing newborn. In this case, the RR is correctly estimated in 460 out of 528 temporal windows (i.e., with a probability of correct estimation equal to 0.871), with a RMSE equal to 0.083 Hz, which corresponds to a normalized RMSE of 9.7%.

## 5.5 Conclusion

Following the presentation of the statistical models of respiratory patterns and breathing diseases in Chapter 4, in this chapter suitable simulators have been introduced. More precisely, software- and hardware-based simulators have been described, both driven by CTMC-based statistical models defined in the previous Chapter 4.

The first simulator has the capability to manipulate video streams of a regularly breathing patient in order to insert apnea events. An extension of this simulator has then been developed, which is able to also modify the breathing pattern of the patient, driven by a multiple-state CTMC.

The second one, composed by a manikin of a newborn with suitable moving parts, is a "physical" simulator able to reproduce realistic breathing movements.

Both simulators, are then employed to create a data set of video recordings to assess the performance of video processing-based algorithms developed in Chapter 3. The results confirm the effectiveness of presented algorithms for the estimation of the RR and the detection of apnea events.

# Conclusions

In this dissertation, various algorithms for remote monitoring and analysis of movements related to respiration were developed. Information extracted by video processing-based algorithms are then employed: first, for the detection of specific breathing disorders, in particular apneas, defined as temporary absence of respiration, and then for the estimation of the Respiratory Rate (RR). The devised algorithms were tested and calibrated for the monitoring of newborns, who need to be carefully overseen.

Because of complications in revealing subtle motion in video streams, such as respiration movements of infants, the first solution exploits a motion magnification algorithm, also known as Eulerian Video Magnification (EVM), with the purpose to enhance and highlight small periodic movements so that they become detectable. By using this method, it was demonstrated that, applying the EVM algorithm paired with a technique for the detection of periodicity based on a Maximum Likelihood (ML) analysis, possible lack of breathing movements can be detected and a prompt warning signal to safeguard the patient's health.

The motion magnification technique was then improved, in order to embed it with the algorithm for the extraction of motion signal end reinforce the estimation of periodicity. The newly proposed algorithm was tested on regularly breathing newborns to measure the RR estimation performance, obtaining promising results compared with those obtained by the gold standard: the pneumograph.

The previously described techniques make both use of the ML approach for the estimation of the fundamental frequency related with respiration. Exploiting the properties and the advantage of this approach and relying on the fact that breathing involves quasi-periodic movements, a new proposed algorithm employs a generalized model of pixel-wise periodicity and applies the ML criterion directly on raw video streams. The system first selects suitable Regions of Interest (ROI) mainly affected by respiratory movements: the obtained ROI are then jointly analyzed for the estimation of the RR of the framed patient. Furthermore, a large motion detection algorithm is applied in order to exclude the ROI affected by movements unrelated to respiration. Results show that the latter system can improve performance with respect to the previous one, thanks to a better management of the large motion of the patient and environmental changes.

In the second part of this thesis, the problem concerning with the lack of large video databases, necessary to design, implement and test video processing-based algorithms for respiration monitoring, was discussed. As a possible solution to this important issue, a statistical model, based on Continuous-Time Markov Chain (CTMC), of the RR patterns of a patient is developed.

First, a simple two-state model to statistically describe the occurrence of apneas in patients suffering from breathing disorder is introduced. The obtained model is able to reproduce realistic patterns of the succession of regular breathing time intervals and apnea events with a properly selected duration. This model is then extended to a multiple-state CTMC to represent complete and realistic breathing patterns, considering the variation of the RR of a patient over time. The statistical model is driven by real data extracted from monitored patients by a pneumographic device. In this way, the developed models are validated by comparing the statistical characteristics of the real cases with the synthetic ones, extracted by simulations of the CTMCs models.

Following this strategy, suitable simulators, based on the previously described statistical models, to test video-processing based algorithms were developed. In particular, two simulators are described: software- and hardware-based. The first simulator directly manipulates video recordings in order to introduce apnea events or create custom respiration patterns in a regularly breathing patient. The second one consists of a manikin of a newborn with a “moving chest”, able to reproduce respiratory movements following the simulation of the statistical models of breathing patterns. Finally, the two simulators are used to obtain performance evaluation of the previously proposed video processing-based algorithms, showing the effectiveness of both algorithms and the relevant simulators.

Possible future research activity can focus on the implementation of the video processing algorithms in mobile or portable devices, to obtain new products to integrate in eHealth or mHealth environments. The algorithms can also be extended in order to simultaneously monitor other vital signs, such as the Heart Rate (HR). In addition, the statistical models can be further extended to consider the influence of variations of other physiological parameters, working towards more precise models and simulators which should allow the test of new non-invasive and contactless video processing-based methods for the monitoring of respiration.

# List of Publications

The list of publications based on this thesis work at time of preparation of this final version is here reported.

**International Journals**

- **D. Alinovi**, G. Ferrari, F. Pisani, and R. Raheli, "Markov chain modeling and simulation of breathing patterns," *Biomed. Signal Process. Control* (Elsevier), vol. 33, pp. 245–254, Mar. 2017. `doi:10.1016/j.bspc.2016.12.002`.
- L. Cattani, **D. Alinovi**, G. Ferrari, R. Raheli, E. Pavlidis, C. Spagnoli, and F. Pisani, "Monitoring infants by automatic video processing: a unified approach to motion analysis," *Comput. Biol. Med.* (Elsevier), vol. 80, pp. 158–165, Jan. 2017. `doi:10.1016/j.compbiomed.2016.11.010`.

**International Conferences**

- **D. Alinovi**, G. Ferrari, F. Pisani, and R. Raheli, "Respiratory rate monitoring by maximum likelihood video processing," in *Proc. 2016 IEEE Int. Symp. Signal Process. and Inf. Technol.* (ISSPIT), Limassol, Cyprus, Dec. 2016. To be published.
- **D. Alinovi**, L. Cattani, G. Ferrari, F. Pisani, and R. Raheli, "Video simulation of apnoea episodes," in *Proc. 2015 IEEE Int. Conf. Multi-*

*media and Expo Workshops* (ICMEW), Turin, Italy, June 2015, pp. 1-6. `doi:10.1109/ICMEW.2015.7169869`.

- **D. Alinovi**, L. Cattani, G. Ferrari, F. Pisani, and R. Raheli, “Spatio-temporal video processing for respiratory rate estimation,” in *Proc. 2015 IEEE Int. Symp. Med. Meas. and Applicat.* (MeMeA), Turin, Italy, May 2015, pp. 12-17. `doi:10.1109/MeMeA.2015.7145164`.
- L. Cattani, **D. Alinovi**, G. Ferrari, R. Raheli, E. Pavlidis, C. Spagnoli, and F. Pisani, “A wire-free, non-invasive, low-cost video processing-based approach to neonatal apnoea detection,” in *Proc. 2014 IEEE Workshop Biometric Meas. and Syst. Security and Med. Applicat.* (BIOMS), Rome, Italy, Oct. 2014, pp. 67-73. `doi:10.1109/BIOMS.2014.6951538`.

# Bibliography

[1] J. Tu, K. Inthavong, and G. Ahmadi, *Computational Fluid and Particle Dynamics in the Human Respiratory System*, 1st ed. Dordrecht, The Netherlands: Springer, 2013.

[2] M. A. Cretikos, R. Bellomo, K. Hillman, J. Chen, S. Finfer, and A. Flabouris, “Respiratory rate: the neglected vital sign,” *Med. J. Aust.* (AMPCo), vol. 188, no. 11, pp. 657–659, June 2008.

[3] F. Q. Al-Khalidi, R. Saatchi, D. Burke, H. Elphick, and S. Tan, “Respiration rate monitoring methods: A review,” *Pediatr. Pulm.* (Wiley), vol. 46, no. 6, pp. 523–529, June 2011.

[4] C. R. Marcus, “Sleep-disordered breathing in children,” *ATS Am. J. Respir. Crit. Care. Med.*, vol. 164, no. 1, pp. 16–30, July 2008.

[5] J. A. Rocker and J. Israel. (2015, Jan.) Pediatric apnea. Medscape. Emergency Medicine. Accessed on Dec. 7$^{\text{th}}$, 2016. [Online]. Available: http://emedicine.medscape.com/article/800032

[6] M.-M. Nano, “Automated detection of central apnea in preterm infants,” Master’s thesis, Delft University of Technology, Oct. 2015.

[7] T. W. Chin, J. J. Chen, and S. Maharaj. (2014, June) Congenital central hypoventilation syndrome. Medscape. Pediatrics: General Medicine. Accessed on Dec. 7$^{\text{th}}$, 2016. [Online]. Available: http://emedicine.medscape.com/article/1002927.

[8] F. Healy and C. L. Marcus, “Congenital central hypoventilation syndrome in children,” *Paediatr. Respir. Rev.* (Elsevier), vol. 12, no. 4, pp. 253–263, Dec. 2011.

[9] W. H. Spriggs, *Essentials of Polysomnography*, 2nd ed. Burlington, MA, USA: Jones & Bartlett Learning, 2015.

[10] Sleep Apnea Guide. (2016) The polysomnogram test. Accessed on Dec. 4th, 2016. [Online]. Available: http://www.sleep-apnea-guide.com/polysomnogram.html

[11] S. Ansari, K. Najarian, K. Ward, and M. H. Tiba, “Extraction of respiratory rate from impedance signal measured on arm: A portable respiratory rate measurement device,” in *Proc. IEEE Int. Conf. Bioinf. and Biomed.* (BIBM), Washington D.C., USA, Nov. 2009, pp. 197–202.

[12] G.-Z. Liu, D. Wu, Z.-Y. Mei, Q.-S. Zhu, and L. Wang, “Automatic detection of respiratory rate from electrocardiogram, respiration induced plethysmography and 3D acceleration signals,” *J. Cent. South. Univ.* (Springer), vol. 20, no. 9, pp. 2423–2431, Sept. 2013.

[13] H. Lee, C. G. Rusin, D. E. Lake, M. T. Clark, L. Guin, T. J. Smoot, A. O. Paget-Brown, B. D. Vergales, J. Kattwinkel, J. R. Moorman, and J. B. Delos, “A new algorithm for detecting central apnea in neonates,” *IOP Physiol. Meas.*, vol. 33, no. 1, pp. 1–17, Jan. 2012.

[14] A. Othonos, “Fiber Bragg gratings,” *AIP Rev. Sci. Instrum.*, vol. 68, no. 12, pp. 4309–4341, Dec. 1997.

[15] V. Mishra and N. Singh, “Optical fiber gratings in perspective of their applications in biomedicine,” in *Biomedicine*, 1st ed., C. Lin, Ed. Rijeka, Croatia: InTech, Mar. 2012, ch. 6, pp. 125–146.

[16] J. Witt, F. Narbonneau, M. Schukar, K. Krebber, J. D. Jonckheere, M. Jeanne, D. Kinet, B. Paquet, A. Depre, L. T. D’Angelo, T. Thiel,

and R. Logier, “Medical textiles with embedded fiber optic sensors for monitoring of respiratory movement,” *IEEE Sensors J.*, vol. 12, no. 1, pp. 246–254, Jan. 2012.

[17] L. Dziuda, F. W. Skibniewski, M. Krej, and J. Lewandowski, “Monitoring respiration and cardiac activity using fiber Bragg grating-based sensor,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 7, pp. 1934–1942, July 2012.

[18] A. Pantelopoulos and N. G. Bourbakis, “A survey on wearable sensor-based systems for health monitoring and prognosis,” *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 1, pp. 1–12, Jan. 2010.

[19] S. Bouwstra, W. Chen, L. Feijs, and S. B. Oetomo, “Smart jacket design for neonatal monitoring with wearable sensors,” in *Proc. 6th Int. Workshop Wearable and Implantable Body Sensor Netw.*, Berkeley, CA, USA, June 2009, pp. 162–167.

[20] D. Dei, G. Grazzini, G. Luzi, M. Pieraccini, C. Atzeni, S. Boncinelli, G. Camiciottoli, W. Castellani, M. Marsili, and J. Lo Dico, “Non-contact detection of breathing using a microwave sensor,” *Sensors* (MDPI), vol. 9, no. 4, pp. 2574–2585, Apr. 2009.

[21] S. Suzuki, T. Matsui, H. Kawahara, H. Ichiki, J. Shimizu, Y. Kondo, S. Gotoh, H. Yura, B. Takase, and M. Ishihara, “A non-contact vital sign monitoring system for ambulances using dual-frequency microwave radars,” *Med. Biol. Eng. Comput.* (Springer), vol. 47, no. 1, pp. 101–105, Jan. 2009.

[22] S. D. Min, J. K. Kim, H. S. Shin, Y. H. Yun, C. K. Lee, and J.-H. Lee, “Noncontact respiration rate measurement system using an ultrasonic proximity sensor,” *IEEE Sensors J.*, vol. 10, no. 11, pp. 1732–1739, Nov. 2010.

[23] N. Patwari, J. Wilson, S. Ananthanarayanan, S. Kasera, and D. Westenskow, “Monitoring breathing via signal strength in wireless networks,” *IEEE Trans. Mobile Comput.*, vol. 13, no. 8, pp. 1774–1786, Aug. 2014.

[24] J. J. Volpe, *Neurology of the Newborn*, 5th ed. Philadelphia, PA, USA: Saunders Elsevier, 2008.

[25] N. B. Karayiannis, S. Srinivasan, R. Bhattacharya, M. S. Wise, J. D. Frost, and E. M. Mizrahi, “Extraction of motion strength and motor activity signals from video recordings of neonatal seizures,” *IEEE Trans. Med. Imag.*, vol. 20, no. 9, pp. 965–980, Sept. 2001.

[26] N. B. Karayiannis, A. Sami, J. D. Frost, M. S. Wise, and E. M. Mizrahi, “Automated extraction of temporal motor activity signals from video recordings of neonatal seizures based on adaptive block matching,” *IEEE Trans. Biomed. Eng.*, vol. 52, no. 4, pp. 676–686, Apr. 2005.

[27] N. B. Karayiannis, Y. Xiong, J. D. Frost, M. S. Wise, and E. M. Mizrahi, “Improving the accuracy and reliability of motion tracking methods used for extracting temporal motor activity signals from video recordings of neonatal seizures,” *IEEE Trans. Biomed. Eng.*, vol. 52, no. 4, pp. 747–749, Apr. 2005.

[28] G. M. Kouamou Ntonfo, G. Ferrari, R. Raheli, and F. Pisani, “Low-complexity image processing for real-time detection of neonatal clonic seizures,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 375–382, May 2012.

[29] L. Cattani, G. M. Kouamou Ntonfo, F. Lofino, G. Ferrari, R. Raheli, and F. Pisani, “Maximum-likelihood detection of neonatal clonic seizures by video image processing,” in *Proc. 8th Int. Symp. Med. Inform. and Commun. Technol.* (ISMICT), Florence, Italy, Apr. 2014, pp. 1–5.

[30] L. Cattani, D. Alinovi, G. Ferrari, R. Raheli, E. Pavlidis, C. Spagnoli, and F. Pisani, “Monitoring infants by automatic video processing: a unified

approach to motion analysis," *Comput. Biol. Med.* (Elsevier), 2016, to be published.

[31] J. Jacobs, "Detecting neonatal seizures: A challenge accepted!" *Clin. Neurophysiol.* (Elsevier), vol. 125, no. 8, pp. 1501–1503, Aug. 2014.

[32] F. Pisani, E. Pavlidis, L. Cattani, G. Ferrari, R. Raheli, and C. Spagnoli, "Optimizing detection rate and characterization of subtle paroxysmal neonatal abnormal facial movements with multi-camera video-electroencephalogram recordings," *Neuropediatrics* (Elsevier), vol. 47, no. 3, pp. 169–174, June 2016.

[33] J. Webb and J. Ashley, *Beginning Kinect programming with the Microsoft Kinect SDK*, 1st ed. New York, NY, USA: Apress, 2012.

[34] Y.-W. Bai, W.-T. Li, and Y.-W. Chen, "Design and implementation of an embedded monitor system for detection of a patient's breath by double webcams," in *Proc. IEEE Int. Symp. Med. Meas. and Applicat.* (MeMeA), Ottawa, Canada, Apr. 2010, pp. 171–176.

[35] K. S. Tan, R. Saatchi, H. Elphick, and D. Burke, "Real-time vision based respiration monitoring system," in *Proc. 7th Int. Symp. Commun. Syst. Netw. and Digit. Signal Process.* (CSNDSP), Newcastle upon Tyne, UK, July 2010, pp. 770–774.

[36] M. C. Yu, J. L. Liou, S. W. Kuo, M. S. Lee, and Y. P. Hung, "Noncontact respiratory measurement of volume change using depth camera," in *Proc. Annu. Int. Conf. IEEE Eng. Med. and Biol. Soc.* (EMBC), San Diego, CA, USA, Aug. 2012, pp. 2371–2374.

[37] C. G. Scully, J. Lee, J. Meyer, A. M. Gorbach, D. Granquist-Fraser, Y. Mendelson, and K. H. Chon, "Physiological parameter monitoring from optical recordings with a mobile phone," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 2, pp. 303–306, Feb. 2012.

[38] M. Bartula, T. Tigges, and J. Muehlsteff, "Camera-based system for contactless monitoring of respiration," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. and Biol. Soc.* (EMBC), Osaka, Japan, July 2013, pp. 2672–2675.

[39] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, "Multiparameter respiratory rate estimation from the photoplethysmogram," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 7, pp. 1946–1953, July 2013.

[40] C.-W. Wang, A. Hunter, N. Gravill, and S. Matusiewicz, "Unconstrained video monitoring of breathing behavior and application to diagnosis of sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 396–404, Feb. 2014.

[41] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. W. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *IOP Physiol. Meas.*, vol. 35, no. 5, pp. 807–831, May 2014.

[42] R. Janssen, W. Wang, A. Moço, and G. de Haan, "Video-based respiration monitoring with automatic region of interest detection," *IOP Physiol. Meas.*, vol. 37, no. 1, pp. 100–114, Jan. 2016.

[43] O. Marques, *Practical Image and Video Processing using MATLAB*, 1st ed., IEEE Press, Ed. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2011.

[44] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Pearson - Prentice Hall, 2008.

[45] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 65:1 – 65:8, July 2012.

[46] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Upper Saddle River, NJ, USA: Prentice Hall, 1993, vol. 1.

[47] P. J. Burt and E. H. Adelson, “The Laplacian pyramid as a compact image code,” *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.

[48] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd ed., ser. Prentice Hall Signal Processing. Upper Saddle River, NJ, USA: Pearson - Prentice Hall, 2010.

[49] L. B. Jackson, *Digital Filters and Signal Processing*, 3rd ed. Norwell, MA, USA: Kluwer Academic Publishers - Springer, 1996.

[50] M. Richardson, “The respiratory system - part 4: breathing,” *Nurs. Times* (EMAP), vol. 102, no. 24, pp. 26–27, June 2006.

[51] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1996.

[52] R. Keys, “Cubic convolution interpolation for digital image processing,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981.

[53] D. Alinovi, G. Ferrari, F. Pisani, and R. Raheli, “Markov chain modeling and simulation of breathing patterns,” *Biomed. Signal Process. Control* (Elsevier), 2016, to be published. Available on arXiv.org [arXiv:1610.01444].

[54] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York, NY, USA: McGraw-Hill, 2002.

[55] J. R. Norris, *Markov Chains*, 1st ed. Cambridge, UK: Cambridge University Press, 1997.

[56] A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd ed. Upper Saddle River, NJ, USA: Pearson Prentice Hall, 2008.

[57] K. A. Wyka, P. J. Mathews, and J. Rutkowski, *Foundations of Respiratory Care*, 2nd ed. Clifton Park, NY, USA: Cengage Learning, 2011.

[58] J. Max, "Quantizing for minimum distortion," *IEEE Trans. Inf. Theory*, vol. 6, no. 1, pp. 7–12, Mar. 1960.

[59] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[60] P. Metzner, E. Dittmer, T. Jahnke, and C. Schütte, "Generator estimation of Markov jump processes," *J. Comput. Physics* (Elsevier), vol. 227, no. 1, pp. 353–375, Nov. 2007.

[61] M. Bladt and M. Sørensen, "Statistical inference for discretely observed Markov jump processes," *J. Royal Statistical Soc. Series B (Statistical Methodology)* (Wiley), vol. 67, no. 3, pp. 395–410, 2005.

[62] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: John Wiley & Sons, 2006.

[63] E. Promayon, P. Baconnier, and C. Puech, *Physically-based model for simulating the human trunk respiration movements*, ser. Lect. Notes Comput. Sci. Berlin, Heidelberg, Germany: Springer, Mar. 1997, vol. 1205, pp. 379–388.

[64] V. B. Zordan, B. Celly, B. Chiu, and P. C. DiLorenzo, "Breathe easy: model and control of human respiration for computer animation," *Graph. Models* (Elsevier), vol. 68, no. 2, pp. 113–132, Mar. 2006.

[65] S. M. Cervera, D. Nikoli, A. Barney, and R. Allen, "A model of breathing abnormalities in sleep for development of classification and diagnosis techniques," in *Proc. 3rd Int. Symp. Appl. Sci. Biomed. and Commun. Technol.* (ISABEL), Rome, Italy, Nov. 2010, pp. 1–2.

[66] F. Zhen-Ping and B. Kang, “Study of flame simulation based on dynamic video texture,” in *Proc. 7th Int. Conf. Natural Comput.* (ICNC), Shanghai, China, July 2011, pp. 1863–1866.

[67] Y.-Y. Chuang, D. B. Goldman, K. C. Zheng, B. Curless, D. H. Salesin, and R. Szeliski, “Animating pictures with stochastic motion textures,” *ACM Trans. Graph.*, vol. 24, no. 3, pp. 853–860, July 2005.

[68] R. E. Carlson and F. N. Fritsch, “Monotone piecewise cubic interpolation,” *SIAM J. Numer. Anal.*, vol. 22, no. 2, pp. 386–400, Apr. 1985.

[69] G. D. Knott, *Interpolating Cubic Splines*, 1st ed. Basel, Switzerland: Birkhäuser Basel, 2000.

[70] J. Blum, *Exploring Arduino: Tools and Techniques for Engineering Wizardry*, 1st ed. Indianapolis, IN, USA: John Wiley & Sons, 2013.

[71] L. Cattani, D. Alinovi, G. Ferrari, R. Raheli, E. Pavlidis, C. Spagnoli, and F. Pisani, “A wire-free, non-invasive, low-cost, video processing-based approach to neonatal apnoea detection,” in *Proc. IEEE Workshop Biometric Meas. and Syst. Security and Med. Applicat.* (BIOMS), Rome, Italy, Oct. 2014, pp. 67–73.

[72] D. Alinovi, L. Cattani, G. Ferrari, F. Pisani, and R. Raheli, “Spatio-temporal video processing for respiratory rate estimation,” in *Proc. IEEE Int. Symp. Med. Meas. and Applicat.* (MeMeA), Turin, Italy, June 2015, pp. 12–17.

[73] A. G. Lalkhen and A. McCluskey, “Clinical tests: sensitivity and specificity,” *Contin. Educ. Anaesth. Crit. Care & Pain* (Oxford University Press), vol. 8, no. 6, pp. 221–223, Dec. 2008.

[74] A. S. Glas, J. G. Lijmer, M. H. Prins, G. J. Bonsel, and P. M. M. Bossuyt, “The diagnostic odds ratio: a single indicator of test performance,” *J. Clin. Epidemiol.* (Elsevier), vol. 56, no. 11, pp. 1129–1135, Nov. 2003.

[75] J. A. Swets, “Measuring the accuracy of diagnostic systems,” *Science* (AAAS), vol. 240, no. 4857, pp. 1285–1293, June 1988.

# Acknowledgments

First and foremost, I wish to sincerely thank my supervisor, Prof. Riccardo Raheli. With his wisdom, he continuously guided me during the years of Ph.D. course, inspired the research activity and helped me when I thought I shouldn't succeed.

I would like to thank two other people with whom I have closely worked: my tutor, Prof. Gianluigi Ferrari, for his willingness and the insightful discussions, and Prof. Francesco Pisani, affiliated to the Department of Neuroscience (University of Parma), for his cordiality, the feedback and help provided in medical stuff.

I also thank all friends and colleagues encountered in the Department of Information Engineering (University of Parma), especially old and current members of the Multimedia Lab and IoT Lab (previously WASN Lab): thank you for the good times together.

Finally, but not least, I'm grateful to my parents and family for these years of support, especially in though times; they always encouraged me during this "crazy" intent to undertake doctoral studies.