Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

# Motion magnification algorithms for video-based breathing monitoring

Veronica Mattioli [a,*], Davide Alinovi [1], Gianluigi Ferrari [a], Francesco Pisani [b], Riccardo Raheli [a]

[a] Department of Engineering and Architecture, University of Parma, Parco Area delle Scienze 181/A, 43124 Parma, Italy
[b] Department of Human Neuroscience, Sapienza University of Rome, Viale dell'Università 30, 108, 00185 Rome, Italy

ARTICLE INFO

*Keywords:*
Respiratory rate estimation
Motion magnification
Maximum likelihood
Video processing

ABSTRACT

In this paper, we present two video processing techniques for contact-less estimation of the Respiratory Rate (RR) of framed subjects. Due to the modest extent of movements related to respiration in both infants and adults, specific algorithms to efficiently detect breathing are needed. For this reason, motion-related variations in video signals are exploited to identify respiration of the monitored patient and simultaneously estimate the RR over time. Our estimation methods rely on two motion magnification algorithms that are exploited to enhance the subtle respiration-related movements. In particular, amplitude- and phase-based algorithms for motion magnification are considered to extract reliable motion signals. The proposed estimation systems perform both spatial decomposition of the video frames combined with proper temporal filtering to extract breathing information. After periodic (or quasi-periodic) respiratory signals are extracted and jointly analysed, we apply the Maximum Likelihood (ML) criterion to estimate the fundamental frequency, corresponding to the RR. The performance of the presented methods is first assessed by comparison with reference data. Videos framing different subjects, i.e., newborns and adults, are tested. Finally, the RR estimation accuracy of both methods is measured in terms of normalized Root Mean Squared Error (RMSE), demonstrating the superiority, performance-wise, of the phase-based method.

## 1. Introduction

Breathing monitoring is a fundamental diagnostic tool to assess the physiological status of a patient. In particular, the Respiratory Rate (RR) is a relevant indicator of potential human respiratory system dysfunctions that may be caused by critical medical conditions. Typical values of the RR in healthy adults at rest lie between 12 and 20 breaths per minute and may vary with age. The RR in newborns and children is usually higher. Abnormal values of the RR may be a sign of serious problems arising from respiratory disorders or complications. For instance, diseases such as chronic obstructive pulmonary disease, asthma, anaemia and epileptic seizures may cause oxygen levels in the blood to significantly drop, potentially leading to cyanosis, cerebral palsy or cardiac arrest and ischaemic events [1].

A constant and careful monitoring of the respiration of a patient is crucial for early diagnosis and intervention that may be lifesaver in some cases. Recent and extensive reviews about current RR monitoring methodologies can be found in [1,2]. In particular, these monitoring systems are typically classified into two main categories: contact-based and contact-less. In [1], a thorough analysis of both categories is presented, whereas [2] focuses on contact-less methods only, which are gaining increasing attention thanks to the advantages they may provide. In fact, they are particularly suitable for remote monitoring, that has become fundamental especially in the pandemic era in which patients affected by COVID-19 need constant medical attention [3]. Furthermore, contact-less devices include Red, Green and Blue (RGB) and Infra Red (IR) cameras, as well as microphones and radars, among other sensors [2], whose costs are significantly lower than those of sophisticated equipment usually deployed in hospital environments. These instruments are also non-invasive, hence more comfortable, as they do not require a direct contact with the body of the patient. On the other hand, contact-based methods include more invasive procedures, such as pneumography [4] and phlebotomy [5]. The former technique allows to measure the thoracic movements by means of an elastic belt placed around the chest of the patient, whereas the latter allows to sample arterial, capillary or venous blood gas. Despite its high accuracy, phlebotomy may be painful and difficult to perform, especially in children and newborns, and may lead to complications such as thrombosis, haemorrhage and aneurysm formation [5].

Other conventional probes to monitor the cardiac and pulmonary activity are the ElectroCardioGram (ECG) and the Pneumogram, which

[*] Corresponding author.
*E-mail addresses:* veronica.mattioli@unipr.it (V. Mattioli), gianluigi.ferrari@unipr.it (G. Ferrari), francesco.pisani@uniroma1.it (F. Pisani), riccardo.raheli@unipr.it (R. Raheli).
[1] Deceased on 16 September 2020.

https://doi.org/10.1016/j.bspc.2023.105148
Received 24 November 2022; Received in revised form 6 May 2023; Accepted 8 June 2023

require wired electrodes to be directly attached to the chest of the patient. The main limitation of these instruments is their deployment, as it is mainly limited to clinical settings and is not suitable for home care. As another example of contact-based devices, we mention the pulse oximeter, that has become very popular nowadays as it allows to easily measure the oxygen saturation in the blood, also in domestic environments. For example, this is, indeed, a very informative parameter about the severity of the COVID-19 disease. The pulse oximeter is usually clipped to the fingertip of a patient and measures the changes in the transmission or reflection of the light emitted by a Light-Emitting Diode (LED) hitting the skin of the subject. This operational principle is referred to as PhotoPlethysmoGraphy (PPG) and has also inspired some contact-less video-based monitoring methodologies, as discussed in the following.

Among contact-less methodologies for the RR monitoring, video processing systems based on IR or RGB sensors are becoming very appealing. For instance, thermal imaging techniques may be exploited to detect temperature variations around the nostrils, as in [6,7], to discriminate between the inhalation and exhalation phases. Despite being robust against illumination changes, thermal imaging techniques are subject to some limitations, being sensitive to the ambient temperature variations and requiring the nasal area to be clearly visible. More robust methods could be obtained by employing RGB cameras for which three main approaches can be identified on the basis of (a) PPG, (b) optical flow, and (c) motion magnification [2].

In [8], the respiratory signal is extracted from selected PPG signals computed on a Region Of Interest (ROI) that surrounds the pit of the neck of the subject. The RR is estimated in frequency and time domains and the performance of different RGB camera sensors is analysed. The PPG principle is also exploited in [9], where the hue channel of the Hue, Saturation and Value (HSV) colour space is considered for the analysis. On the other hand, the optical flow may be exploited to detect and track breathing-related movements, as in [10,11]. However, since respiration movements may be subtle and difficult to detect, especially in newborns, motion magnification techniques may be applied to enhance them. Preliminary attempts may be found, e.g., in [12–15], where the usefulness of these techniques towards more accurate RR measurements was initially demonstrated.

An effective mathematical model of the RR and its possible disorders is presented in [16]. This model is based on a time-continuous Markov chain and enables the implementation of video-based simulations of breathing disorders, which may be useful in the design of RR estimation algorithms.

This paper analyses motion magnification algorithms for RR estimation. Amplitude- and phase-based techniques, respectively inspired by the works in [17,18], are considered. In particular, in [17] spatial and temporal processing is combined to amplify the variations of the pixel intensities for frequency bands of interest obtained by a Laplacian decomposition [19]. In [18], an approximation of the Riesz transform is proposed to perform phase amplification of motion signal. This paper improves upon the approaches in these references by demonstrating that video reconstruction with amplified motion is not necessary in breathing monitoring, which can instead be directly performed on the amplified motion signal components, thus improving both efficiency and effectiveness. Once the amplified motion signal components are obtained, an estimation technique, based on the Maximum Likelihood (ML) principle [20], can, indeed, be applied to estimate the RR.

Related work that may be worth mentioning is the following. The paper [13] is based on the method proposed in [17], but uses a wavelet pyramid decomposition to obtain the frequency bands of interest. Amplitude magnification and the unnecessary final video reconstruction are performed, as also done in [12]. The work in [15], which we were referred to during the review process, is also based on the magnification method in [17], but frequency bands are obtained by applying the Hermite transform to each video frame and a Convolutional Neural Network (CNN) is trained to classify the presence of inhalation or exhalation within the processed frames. The work in [14], based on the phase magnification technique described in [18], presents a method for video stabilization for handheld cameras and the procedure to extract useful phase signals in this context. These works represent significant contributions as initial demonstrations of the effectiveness of motion amplification techniques applied to the task of RR estimation. Therefore, they motivate further investigation.

The main contributions of this paper with respect to the state of the art are the following:

- The RR estimation is directly performed on the magnified signals, rather than on the reconstructed video sequence
- The employed estimation technique, i.e., the ML principle, represents a novelty in the context of video processing for the extraction of periodic features, such as the RR
- A unified and comprehensive comparative analysis of the amplitude- and phase-based methods is presented
- An extensive performance analysis is carried out in realistic experimental scenarios
- The performance advantage of the phase-based method is demonstrated.

This paper expands upon preliminary conference versions by some of the authors [21–23].

The remainder of the paper is organized as follows. In Section 2, the extraction of the motion signals is described and two techniques for motion magnification are detailed. In Section 3, the RR estimation method is presented and the procedure of automatic selection of the ROIs with breathing-related movements is described, along with a decision strategy to discard unsuitable ROIs. In Section 4, the performance of the considered methods is discussed and compared against reference data. Finally, in Section 5 conclusions are drawn.

## 2. Motion signal extraction

In this section, the extraction of motion signals from video sequences is described. Full-frame video sequences are initially considered as inputs to the proposed methods for an effective and simple description. Nevertheless, ROIs can also be extracted to reduce the computational complexity as well as to improve robustness. To this end, a method to automatically select ROIs will be presented in Section 3.1 and a method to discard unsuitable ROIs, based on the detection of large movements unrelated with respiration, will be presented in Section 3.2.

In the following, we refer to a generic gray scale video sequence, acquired with a sampling rate $f_s$ (dimension: [frame/s]), as a discrete signal $f[\mathbf{u}, n]$ that defines the pixel intensities at position $\mathbf{u} = (u_1, u_2)$ at the $n$-th frame. Each frame has size $U_1 \times U_2$ (dimension: [pixel × pixel]) and is sampled at time instants $nT_s$ (dimension: [s]), where $T_s = 1/f_s$ is the sampling period (dimension: [s]). The videos considered in this paper, recorded by RGB cameras, can be converted to gray scale [24].

### 2.1. Amplitude-based motion magnification

Amplitude-based techniques for motion magnification aim at linearly amplifying variations of each pixel intensity over time. The method proposed in [17], called Eulerian Video Magnification (EVM), performs temporal processing on different spatial frequency bands obtained by decomposing each frame of the input video into a set of subimages. The processed and unprocessed video subsignals are finally recombined to obtain the amplified output video. Unlike the preliminary work in [12], we present here a spatio-temporal approach to extract useful motion signals, inspired by the EVM algorithm in [17], in which this final recombination step is not performed because not of interest for the purpose of breathing monitoring. An illustrative overview of the proposed method is shown in Fig. 1, where each processing step is associated with a diagram block and is detailed hereafter.

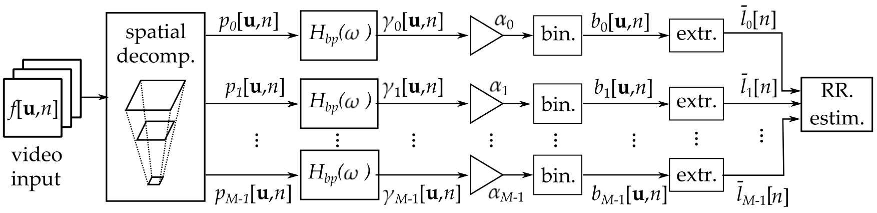


**Fig. 1.** Amplitude-based (spatio-temporal) RR estimation algorithm.

*Spatial decomposition.* As a first step, each frame of the video $f[\mathbf{u},n]$ is decomposed into a set of $M$ subimages with scaled resolutions, each representing a different spatial frequency band. This approach is known as multi-scale decomposition. The obtained $M$ scaled subimages are referred to as "levels" and are sorted according to a decreasing resolution criterion. The multi-scale image decomposition is performed here by computing a Laplacian pyramid [19] of the input frame, as described in the following. First, a Gaussian pyramid [19] is derived, where $g_0[\mathbf{u},n] = f[\mathbf{u},n]$ is set as the bottom level, that corresponds to the highest spatial frequency band and is characterized by the highest resolution. Upper levels, representing lower spatial frequency bands and characterized by lower resolutions, are recursively computed according to a "reduce" function defined as

$$g_m[\mathbf{u},n] = \sum_{k_1=-R_M}^{+R_M} \sum_{k_2=-R_M}^{+R_M} w[k_1,k_2] g_{m-1}[2u_1-k_1, 2u_2-k_2, n] \tag{1}$$

where $m = 1, \dots M-1$ denotes the $m$-th pyramid level, $w[k_1,k_2]$ is a proper truncated Gaussian low-pass filter, designed according to specific constraints described in [19], and $R_M$ is a positive integer that specifies the size of this filter as $(2R_M+1)\times(2R_M+1)$. An "expand" function can also be defined as

$$\hat{g}_m[\mathbf{u},n] = 4 \sum_{k_1=-R_M}^{+R_M} \sum_{k_2=-R_M}^{+R_M} w[k_1,k_2] g_{m+1}\left[\frac{u_1-k_1}{2}, \frac{u_2-k_2}{2}, n\right] \tag{2}$$

to obtain a specific level by expanding the dimensions of the upper one (with lower resolution) by interpolation. The filter mask $w[k_1,k_2]$ is the same in (1) and (2).

The Laplacian pyramid levels are derived from (1) and (2) as

$$p_m[\mathbf{u},n] = \begin{cases} g_m[\mathbf{u},n] - \hat{g}_m[\mathbf{u},n] & m = 1,\dots,M-2 \\ g_m[\mathbf{u},n] & m = M-1 \end{cases} \tag{3}$$

where $p_{M-1}[\mathbf{u},n] = g_{M-1}[\mathbf{u},n]$ is set as the highest-index level and describes the lowest spatial frequency band. The expression in (3) represents the error image between a level of the Gaussian pyramid $g_m$ and the same level $\hat{g}_m$ obtained by expanding the upper one according to the function in (2).

The operation of spatial decomposition is highlighted in the first block of the diagram in Fig. 1.

*Temporal filtering.* Once spatial processing is performed and a spatial decomposition based on the Laplacian pyramid is obtained, each level is pixel-wise temporally filtered to extract a frequency band that corresponds to a typical range of RR. A Butterworth filter of the second order with Infinite Impulse Response (IIR) can be selected as a proper temporal digital band-pass filter. Its transfer function can be expressed as

$$H_{bp}(z) = K\frac{(1+z^{-1})(1-z^{-1})}{(1-pz^{-1})(1-p^*z^{-1})} \tag{4}$$

where the scale factor $K$ and the complex conjugates poles $p$ and $p^*$ can be computed following the filter design rules to satisfy the requirements for the lower and upper 3-dB cut-off frequencies $f_{\mathrm{L}}^{\mathrm{co}}$ and $f_{\mathrm{H}}^{\mathrm{co}}$ [25]. In this work, the cut-off frequencies of the filter are set according to the framed subject: for adults $f_{\mathrm{L}}^{\mathrm{co}} = 0.19$ Hz and $f_{\mathrm{H}}^{\mathrm{co}} = 0.9$ Hz, corresponding to a range of $11 \div 54$ breath/min, whereas for newborns $f_{\mathrm{L}}^{\mathrm{co}} = 0.3$ Hz and $f_{\mathrm{H}}^{\mathrm{co}} = 1.1$ Hz, corresponding to a range of $18 \div 66$ breath/min. The frequency response of the IIR filter employed for adults is shown in Fig. 2.

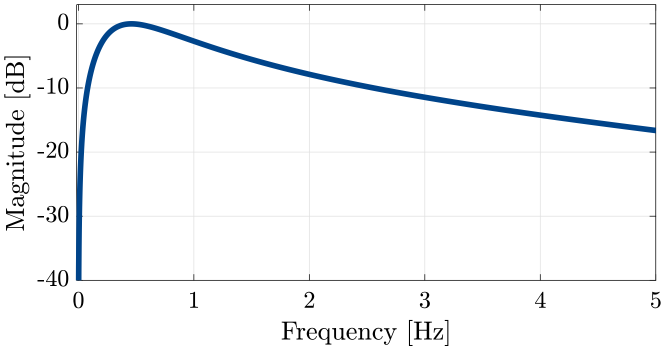


**Fig. 2.** Frequency response of the IIR filter employed for adults.

The temporal processing is represented as a filter bank in Fig. 1 and the obtained filtered levels are denoted as $\{\gamma_m[\mathbf{u},n]\}_{m=0}^{M-1}$.

*Signal amplification.* Each filtered level $\gamma_m[\mathbf{u},n]$, $m = 0,\dots,M-1$, is multiplied by a proper amplification factor to linearly amplify motions related to respiration. The amplification coefficients are denoted as $\{\alpha_m\}_{m=0}^{M-1}$ in Fig. 1 and are properly set according to [17] to avoid noise amplification or motion artefacts. The amplification coefficient for the lowest-index level is set as $\alpha_0 = 1$ and increasing values of amplification are used for higher-index levels, up to $\alpha_{M-2} = 12$. As the highest-index level resolution is too low to provide useful information, $\alpha_{M-1}$ is set to 0.

*Binarization.* Binarization is performed pixel-wise on the amplified signals $\{\gamma_m[\mathbf{u},n]\alpha_m\}_{m=0}^{M-1}$ to reduce the computational complexity (blocks labelled "bin". in Fig. 1). This operation allows to highlight the respiration movements by setting to 1 the pixel intensity values larger than a preset threshold, whereas the rest of the framed scene is set to 0. This operation yields the following binarized levels, also highlighted in Fig. 1

$$b_m[\mathbf{u},n] = \begin{cases} 1 & \text{if } |\{\gamma_m[\mathbf{u},n]\alpha_m\}| \geq \Gamma_{\mathrm{th}} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $\Gamma_{\mathrm{th}}$ is a proper binarization threshold heuristically set to adjust the sensitivity to motion.[2]

*Signal extraction.* As a last step, the motion signals are extracted by spatially averaging each binarized level of the pyramid as

$$\bar{l}_m[n] = \frac{1}{c_m r_m} \sum_{u_1=1}^{c_m} \sum_{u_2=1}^{r_m} b_m[\mathbf{u},n] \tag{6}$$

where $\{c_m\}_{m=0}^{M-1}$ and $\{r_m\}_{m=0}^{M-1}$ are the widths and heights of the binarized frames (blocks labelled "extr". in Fig. 1).

[2] The set of coefficients $\{\alpha_m\}_{m=0}^{M-1}$ and the threshold $\Gamma_{\mathrm{th}}$ in (5) can be scaled by a common factor without affecting the binarized frames. This underdetermined feature can be overcome by setting $\alpha_0 = 1$.

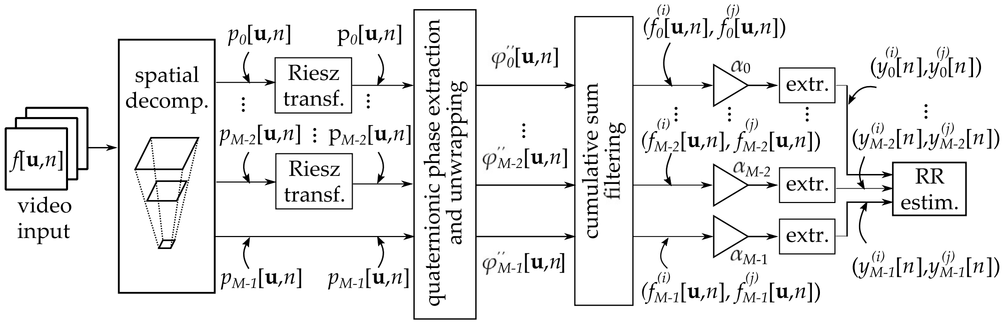

$f[\mathbf{u},n]$
video input
spatial decomp.
$p_0[\mathbf{u},n]$
$\mathrm{p}_0[\mathbf{u},n]$
Riesz transf.
$p_{M\text{-}2}[\mathbf{u},n]$
$\mathrm{p}_{M\text{-}2}[\mathbf{u},n]$
Riesz transf.
$p_{M\text{-}1}[\mathbf{u},n]$
$p_{M\text{-}1}[\mathbf{u},n]$
quaternionic phase extraction and unwrapping
$\varphi''_0[\mathbf{u},n]$
$\varphi''_{M\text{-}2}[\mathbf{u},n]$
$\varphi''_{M\text{-}1}[\mathbf{u},n]$
cumulative sum filtering
$(f^{(i)}_0[\mathbf{u},n], f^{(j)}_0[\mathbf{u},n])$
$(f^{(i)}_{M\text{-}2}[\mathbf{u},n], f^{(j)}_{M\text{-}2}[\mathbf{u},n])$
$(f^{(i)}_{M\text{-}1}[\mathbf{u},n], f^{(j)}_{M\text{-}1}[\mathbf{u},n])$
$\alpha_0$
$\alpha_{M\text{-}2}$
$\alpha_{M\text{-}1}$
extr.
extr.
extr.
$(y^{(i)}_0[n], y^{(j)}_0[n])$
$(y^{(i)}_{M\text{-}2}[n], y^{(j)}_{M\text{-}2}[n])$
$(y^{(i)}_{M\text{-}1}[n], y^{(j)}_{M\text{-}1}[n])$
RR estim.

**Fig. 3.** Phase-based RR estimation algorithm.

### 2.2. Phase-based motion magnification

Amplitude-based motion magnification presents some limitations directly linked to the linear amplification operation. When the analysed motion is small, a pixel intensity variation can be approximated by a first-order Taylor series expansion as described in [26]. If the small motion condition is not verified, or the amplification factor $\alpha_m$ is too large, the approximation is not accurate and the magnification may cause undesired artefacts. Furthermore, for $\alpha_m > 1$, noise is also amplified.

A solution to overcome problems related to linear amplification is provided by phase-based magnification methods, that aim at amplifying the phase of each pyramidal subsignal [18]. We present here an algorithm for motion magnification inspired by [18], whose illustrative overview is shown in Fig. 3, in which each processing step is associated with a diagram block and will be detailed hereafter.

*Spatial decomposition.* Similarly to the amplitude-based (spatio-temporal) method presented in Section 2.1, the first step to extract amplified motion signals, as also highlighted in the first block of Fig. 3, consists of the spatial multi-scale decomposition of each frame of the input video sequence $f[\mathbf{u},n]$. A set of $M$ scaled levels is obtained by computing the Laplacian pyramid [19] according to (1)–(3). An efficient representation of the signals, where amplitudes and phases are highlighted, can now be adopted by computing the Riesz transform [27] of all the pyramid levels $\{p_m[\mathbf{u},n]\}_{m=0}^{M-1}$. The Riesz transform can be defined as a two-Dimensional (2D) generalization of the Hilbert transform and its 2D frequency response in the Fourier domain can be expressed as [28]

$$H(\boldsymbol{\omega}) = \begin{pmatrix} H_1(\omega_1) \\ H_2(\omega_2) \end{pmatrix} = \begin{pmatrix} -j\omega_1/\|\boldsymbol{\omega}\| \\ -j\omega_2/\|\boldsymbol{\omega}\| \end{pmatrix} \tag{7}$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2)$ is the 2D vector of normalized angular frequencies and $\|\cdot\|$ is the euclidean norm operator. The following operation, shown in the second bank of blocks in Fig. 3, is hence performed:

$$\mathcal{R}\{p_m[\mathbf{u},n]\} = \begin{pmatrix} r_{1,m}[\mathbf{u},n] \\ r_{2,m}[\mathbf{u},n] \end{pmatrix} = \begin{pmatrix} h_1[\mathbf{u}] * p_m[\mathbf{u},n] \\ h_2[\mathbf{u}] * p_m[\mathbf{u},n] \end{pmatrix} \tag{8}$$

where $\mathcal{R}\{\cdot\}$ represents the Riesz transform operator, $h_i[\mathbf{u}] = \mathcal{F}^{-1}(H_i(\boldsymbol{\omega}))$, $i = 1, 2$; $*$ denotes the 2D convolution operator, and $\mathcal{F}^{-1}(\cdot)$ is the inverse 2D Fourier transform operator.

The following triple of elements

$$\mathrm{p}_m[\mathbf{u},n] = (p_m[\mathbf{u},n], r_{1,m}[\mathbf{u},n], r_{2,m}[\mathbf{u},n]) \tag{9}$$

is known as the monogenic signal of the $m$-th level. In particular, the combination of $\{\mathrm{p}_m[\mathbf{u},n]\}_{m=0}^{M-2}$ with the last level of the Laplacian pyramid $p_{M-1}[\mathbf{u},n]$ forms a Riesz pyramid.

It may be convenient to represent the monogenic signal in (9) as the quaternion [29]

$$\mathrm{q}_m[\mathbf{u},n] = p_m[\mathbf{u},n] + ir_{1,m}[\mathbf{u},n] + jr_{2,m}[\mathbf{u},n] + k\cdot 0 \tag{10}$$

where $i$, $j$, and $k$ are the imaginary units. Following the quaternionic algebra in [29], the norm and the natural logarithm of the quaternion in (10) can be, respectively, defined as

$$\|\mathrm{q}_m\| = \sqrt{p_m[\mathbf{u},n]^2 + r_{1,m}[\mathbf{u},n]^2 + r_{2,m}[\mathbf{u},n]^2} \tag{11}$$

$$\log(\mathrm{q}_m) = \log(\|\mathrm{q}_m\|) + \frac{ir_{1,m}[\mathbf{u},n] + jr_{2,m}[\mathbf{u},n]}{\|ir_{1,m}[\mathbf{u},n] + jr_{2,m}[\mathbf{u},n]\|} \arccos\frac{p_m[\mathbf{u},n]}{\|p_m[\mathbf{u},n]\|}. \tag{12}$$

The amplitude and the quaternionic phase of (10) can now be computed as

$$A_m[\mathbf{u},n] = \|\mathrm{q}_m[\mathbf{u},n]\| \tag{13}$$

$$i\varphi_m[\mathbf{u},n]\cos(\vartheta_m[\mathbf{u},n]) + j\varphi_m[\mathbf{u},n]\sin(\vartheta_m[\mathbf{u},n]) = \log(\mathrm{q}_m[\mathbf{u},n]/\|\mathrm{q}_m[\mathbf{u},n]\|) \tag{14}$$

where

$$\varphi_m[\mathbf{u},n] = \arctan\left(\left(\sqrt{r_{1,m}[\mathbf{u},n]^2 + r_{2,m}[\mathbf{u},n]^2}\right)/p_m[\mathbf{u},n]\right) \tag{15}$$

$$\vartheta_m[\mathbf{u},n] = \arctan\left(r_{2,m}[\mathbf{u},n]/r_{1,m}[\mathbf{u},n]\right) \tag{16}$$

are the $m$-th phase and orientation, respectively. The main advantage of this signal representation is that the quaternionic phase in (14) is invariant to the signs of the phase and orientation in (15) and (16) [29].

*Temporal filtering.* As a second step of the proposed phase amplification method, temporal filtering is again necessary to select a range of frequencies of interest, as in the amplitude-based approach of Section 2.1. An IIR band-pass second-order Butterworth filter with lower and higher cut-off frequencies $f_{\mathrm{L}}^{\mathrm{co}}$ and $f_{\mathrm{H}}^{\mathrm{co}}$ can be employed to filter the phases of each level of the Riesz pyramid. As discussed in [29], the quaternionic phases in (14) are first unwrapped and their cumulative sum is subsequently filtered. To this purpose, the quaternionic logarithm of the $m$-th ($m = 0, \ldots, M-1$) normalized Riesz pyramid coefficient is computed as

$$\begin{cases} \log(\bar{\mathrm{q}}_m[\mathbf{u},n]) & \text{for } n = 0 \\ \log(\bar{\mathrm{q}}_m[\mathbf{u},n]\bar{\mathrm{q}}_m^{-1}[\mathbf{u},n-1]) & \text{for } n = 1, 2, \ldots \end{cases} \tag{17}$$

where $\bar{\mathrm{q}}_m[\mathbf{u},n] = \frac{\mathrm{q}_m[\mathbf{u},n]}{\|\mathrm{q}_m[\mathbf{u},n]\|}$ is the normalized quaternion [29] and we recall the following definitions of the inverse and conjugate quaternion in (10)

$$\mathrm{q}_m^{-1} = \frac{\mathrm{q}_m^*[\mathbf{u},n]}{\|\mathrm{q}_m\|^2} \tag{18}$$

$$\mathrm{q}_m^* = p_m[\mathbf{u},n] - ir_{1,m}[\mathbf{u},n] - jr_{2,m}[\mathbf{u},n]. \tag{19}$$

Assuming that the orientations are approximately constant in time, the elements in (17) for $n = 1, 2, \ldots$ can be written as

$$i(\varphi'_m[\mathbf{u},n])\cos(\vartheta_m[\mathbf{u}]) + j(\varphi'_m[\mathbf{u},n])\sin(\vartheta_m[\mathbf{u}]) \tag{20}$$

where the term

$$\varphi'_m[\mathbf{u},n] = \varphi_m[\mathbf{u},n] - \varphi_m[\mathbf{u},n-1] \tag{21}$$

is the phase difference. Defining now the unwrapped phase as

$$\varphi''_m[\mathbf{u},n] = \varphi_m[\mathbf{u},0] + \sum_{k=1}^{n} \varphi'_m[\mathbf{u},k] \quad \text{for } n = 1,2,\dots \tag{22}$$

the following cumulative sum can be computed

$$i\varphi''_m[\mathbf{u},n]\cos(\vartheta_m[\mathbf{u}]) + j\varphi''_m[\mathbf{u},n]\sin(\vartheta_m[\mathbf{u}]). \tag{23}$$

By time-filtering the quantity in (23), the following two imaginary quaternionic components are obtained

$$\begin{aligned} f_m^{(i)}[\mathbf{u},n] &= \delta_m[\mathbf{u},n]\cos(\vartheta_m[\mathbf{u}]) \\ f_m^{(j)}[\mathbf{u},n] &= \delta_m[\mathbf{u},n]\sin(\vartheta_m[\mathbf{u}]) \end{aligned} \tag{24}$$

that define the spatial translation due to a framed motion. In Fig. 3, the quaternionic phase extraction and unwrapping operations are associated with a single block that is followed by the cumulative sum filtering block.

*Signal amplification.* Following the approach presented in [30], in order to enhance a motion of interest, the two filtered quaternionic components in (24) at each pyramid level $m \in \{0,\dots,M-1\}$ can be multiplied by the amplification factor $\alpha_m$, $m \in \{0,\dots,M-1\}$, as shown in Fig. 3, obtaining $\{\alpha_m f_m^{(i)}[\mathbf{u},n], \alpha_m f_m^{(j)}[\mathbf{u},n]\}_{m=0}^{M-1}$.

*Signal extraction.* Motion signals can finally be extracted by spatial averaging the amplified and filtered quaternionic components (blocks labelled "extr". in Fig. 3). Considering a frame of size $U_1 \times U_2$, the following signals are obtained

$$\begin{aligned} y_m^{(i)}[n] &= \frac{1}{U_1U_2}\sum_{u_1=1}^{U_1-1}\sum_{u_2=1}^{U_2-1} \alpha_m f_m^{(i)}[\mathbf{u},n] \\ &= \frac{1}{U_1U_2}\sum_{u_1=1}^{U_1-1}\sum_{u_2=1}^{U_2-1} \alpha_m \delta_m[\mathbf{u},n]\cos(\vartheta_m[\mathbf{u}]) \\ y_m^{(j)}[n] &= \frac{1}{U_1U_2}\sum_{u_1=1}^{U_1-1}\sum_{u_2=1}^{U_2-1} \alpha_m f_m^{(j)}[\mathbf{u},n] \\ &= \frac{1}{U_1U_2}\sum_{u_1=1}^{U_1-1}\sum_{u_2=1}^{U_2-1} \alpha_m \delta_m[\mathbf{u},n]\sin(\vartheta_m[\mathbf{u}]). \end{aligned} \tag{25}$$

## 3. Maximum likelihood estimation

Once the motion signals are extracted at each pyramid level, the RR is estimated according to the ML criterion. To this purpose, we first introduce the standard ML principle, that will also be exploited in Section 3.1 to automatically select ROIs in order to focus on areas where the motion is mainly due to breathing.

The ML principle is indeed a reliable and consolidated method that allows to estimate unknown parameters of interest. Since respiration is characterized by quasi-periodic movements of the chest and abdomen, i.e., expansion and relaxation, the ML criterion can be exploited to detect the presence of a fundamental periodic component, corresponding to the RR, and estimate it [22].

The RR estimation operation is embedded in the last blocks of Figs. 1 and 3. As the motion signals are extracted at each pyramid level for both the presented amplitude- and phase-based approaches, a data aggregation method similar to the one proposed in [31] for multiple sensors can be employed.

For the sake of compactness, the motion signals extracted at each pyramid level can be grouped as

$$\mathbf{l}[n] = \begin{bmatrix} \bar{l}_0[n] \\ \bar{l}_1[n] \\ \vdots \\ \bar{l}_{M-1}[n] \end{bmatrix} \tag{26}$$

$$\mathbf{Y}[n] = \begin{bmatrix} y_0^{(i)}[n] & y_0^{(j)}[n] \\ y_1^{(i)}[n] & y_1^{(j)}[n] \\ \vdots & \\ y_{M-1}^{(i)}[n] & y_{M-1}^{(j)}[n] \end{bmatrix} \tag{27}$$

in the case of amplitude (26) and phase (27) components, respectively. Let us define $\mathbf{X}[n]$ as a generic observation model, that can be written in the form of (26) or (27) according to the considered method. The generic size of $\mathbf{X}[n]$ is $M \times C$, where $M$ is the number of considered pyramid levels and the number of columns $C$ is equal to 1 or 2 in the case of (26) or (27), respectively.

Given the nature of the respiration movements of interest, the observation model $\mathbf{X}[n]$ is assumed to have the following structure

$$\mathbf{X}[n] = \mathbf{B} + \mathbf{A}\cos(2\pi f_0 T_s n + \boldsymbol{\Phi}) + \mathbf{W}[n] \tag{28}$$

where $\mathbf{B}$ are the continuous components, $\mathbf{A}$ and $\boldsymbol{\Phi}$ are the amplitudes and phases, respectively, and $\mathbf{W}[n]$ are sequences of independent and identically distributed (i.i.d.) zero-mean Gaussian noise samples, all of size $M \times C$. In (28), the amplitudes $\mathbf{A}$, the fundamental frequency $f_0$, and the phases $\boldsymbol{\Phi}$ are unknown parameters and may be collected as the array of parameters $\boldsymbol{\Theta} = [\mathbf{A}, f_0, \boldsymbol{\Phi}]$. Following the standard method presented in [20, p.193-195] and extending it to the case of multi-dimensional signals, as in [31,32], the parameter array $\boldsymbol{\Theta}$ can be estimated on a window of $N$ frames by minimizing the likelihood function

$$J(\boldsymbol{\Theta}) = \sum_{c=0}^{C-1}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\left[x[m,c,n] - a[m,c]\cos\left(2\pi f_0 T_s n + \phi[m,c]\right)\right]^2 \tag{29}$$

where $x[m,c,n]$, $a[m,c]$ and $\phi[m,c]$ are the generic elements of the matrices $\mathbf{X}[n]$, $\mathbf{A}$ and $\boldsymbol{\Phi}$, respectively. As shown in [20, p.193-195], if the real frequency $f_0$ is not close to 0 or $f_s/2$, an approximate expression of the estimator $\hat{f}_0$ of the fundamental frequency can be derived from (29) as

$$\hat{f}_0 = \underset{f_{\min}\le f\le f_{\max}}{\operatorname{argmax}} \sum_{c=0}^{C-1}\sum_{m=0}^{M-1}\left|\sum_{n=0}^{N-1} x[m,c,n]e^{-j2\pi f T_s n}\right|^2 \tag{30}$$

where the maximization is performed over the limited frequency interval $[f_{\min}, f_{\max}]$, with $f_{\min}$ and $f_{\max}$ being the minimum and the maximum feasible frequencies, respectively, that must be heuristically set.

The amplitudes can similarly be estimated as

$$\hat{a}[m,c] = \frac{2}{N}\sum_{c=0}^{C-1}\sum_{m=0}^{M-1}\left|\sum_{n=0}^{N-1} x[m,c,n]e^{-j2\pi\hat{f}_0 n T_s}\right| \tag{31}$$

and the presence of a significant periodic component is declared, according to [32], only if the following condition is verified

$$\frac{N}{MC}\sum_{c=0}^{C-1}\sum_{m=0}^{M-1}\hat{a}^2[m,c] > \eta \tag{32}$$

where $\eta$ is a properly set threshold.

### 3.1. Region of interest selection

To reduce the computational complexity of the proposed algorithms, a ROI selection algorithm can be exploited to obtain and process video sequences with reduced frame size. In this section, we present an automatic ROI selection algorithm based on the above described ML approach, now applied to the considered video sequence. An illustrative overview of the method is shown in Fig. 4.

Given the generic video sequence $x[\mathbf{u},n]$, the first step for automatically extracting $R$ ROIs consists in selecting $L$ frames where variations are only due to respiration movements. This processing step is associated with the first block of the diagram in Fig. 4. The $L$ frames $\{x[\mathbf{u},n]\}_{n=0}^{L-1}$ are first downsampled in space by an integer value $D$ to reduce the computational complexity, obtaining a new block of frames $\{x_D[\mathbf{u},n]\}_{n=0}^{L-1}$ with a smaller dimension $\lceil U_1/D\rceil \times \lceil U_2/D\rceil$, where $\lceil\cdot\rceil$ represents the ceiling operator. This operation is associated with the second bank of blocks of the diagram in Fig. 4. The ML approach described in Section 3, associated with the third block of Fig. 4, is applied to the downsampled sequences $\{x_D[\mathbf{u},n]\}_{n=0}^{L-1}$ to estimate the

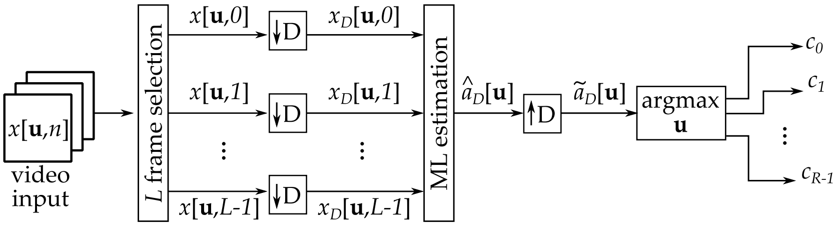


**Fig. 4.** ROI selection algorithm.

fundamental frequency $\hat{f}_0$ and the amplitudes $\hat{a}_D[\mathbf{u}]$, according to (30) and (31), respectively, where $x[m,c,n]$ and $M \times C$ are now replaced by $x_D[\mathbf{u},n]$, i.e., the intensity of the pixel at position $\mathbf{u}$, and $\lceil U_1/D \rceil \times \lceil U_2/D \rceil$, i.e., the size of the frames.

To compute the centres of the selected $R$ ROIs, the matrix of the amplitudes $\hat{a}_D[\mathbf{u}]$, estimated for the reduced frames, is interpolated at the original frame size $U_1 \times U_2$ to estimate the amplitudes $\tilde{a}_D[\mathbf{u}]$ in the original block of frames. The centres $\{c_r\}_{r=0}^{R-1}$ are finally found by selecting the coordinates of the pixels that correspond to the maximum values of $\tilde{a}_D[\mathbf{u}]$. The interpolation and the selection of the ROIs centres are the operations embedded in the fourth and fifth (last) blocks of the diagram in Fig. 4, respectively. This procedure allows to extract $R$ ROIs with a fixed size $W \times W$ and may be repeated over time to deal with changes in the position of the framed subject.

### 3.2. Large motion detection

To discard ROIs where the motion is affected by large movements unrelated with breathing, a further control procedure may be needed. To this purpose, the intensity of the pixel at position $\mathbf{u}$ at the $n$-th frame of the $r$-th ROI can be defined as $x_r[\mathbf{u},n]$ and the pixel-wise difference of consecutive frames can be computed as

$$i[\mathbf{u},n] = x_r[\mathbf{u},n] - x_r[\mathbf{u},n-1]. \tag{33}$$

To reduce the computational complexity, the filtered signal in (33) could also be binarized according to the following binarization rule

$$i_r[\mathbf{u},n] = \begin{cases} 0 & \text{if } \left| x_r[\mathbf{u},n] - x_r[\mathbf{u},n-1] \right| < \gamma_{\text{bin}} \\ 1 & \text{else} \end{cases} \qquad r = 1,2,\dots,R \tag{34}$$

where $\gamma_{\text{bin}}$ is a properly chosen binarization threshold. The average motion signal on the $r$-th region can now be computed as

$$\bar{i}_r[n] = \frac{1}{W^2} \sum_{u_1=1}^{W-1} \sum_{u_2=1}^{W-1} i_r[\mathbf{u},n]. \tag{35}$$

A good decision strategy is such that the $r$-th ROI is discarded if $\bar{i}_r[n]$ in (35) is above a heuristically set threshold, as expressed by the following decision rule:

$$\kappa_r = \begin{cases} 0 & \text{if } \bar{i}_r[n] > \gamma_{\text{th}} \\ 1 & \text{else} \end{cases} \qquad r = 1,2,\dots,R \tag{36}$$

where the binary-valued decision $\kappa_r$ defines the presence ($\kappa_r = 1$) or absence ($\kappa_r = 0$) of large motion inside the $r$-th ROI and $\gamma_{\text{th}}$ is the selected decision threshold.

Finally, the RR is estimated by maximizing the following likelihood function

$$J(\boldsymbol{\Theta}) = \sum_{r=1}^{R} \kappa_r J_r(\boldsymbol{\Theta}) \tag{37}$$

where $J_r(\boldsymbol{\Theta})$ is defined according to (29) and refers to the $r$-th ROI.

The pseudo-code of the proposed phase-based estimation method is detailed in Algorithm 1. The ROI selection algorithm described in Section 3.1 is first applied. The motion signals are extracted for each ROI, as detailed in Section 2, and unsuitable ROIs affected by large motions are discarded, as described in Section 3.2. Finally, the ML estimation algorithm described in Section 3 is applied to obtain the estimated respiratory rate $\hat{f}_0$. The similar pseudo-code of the proposed amplitude-based estimation method is omitted for brevity.

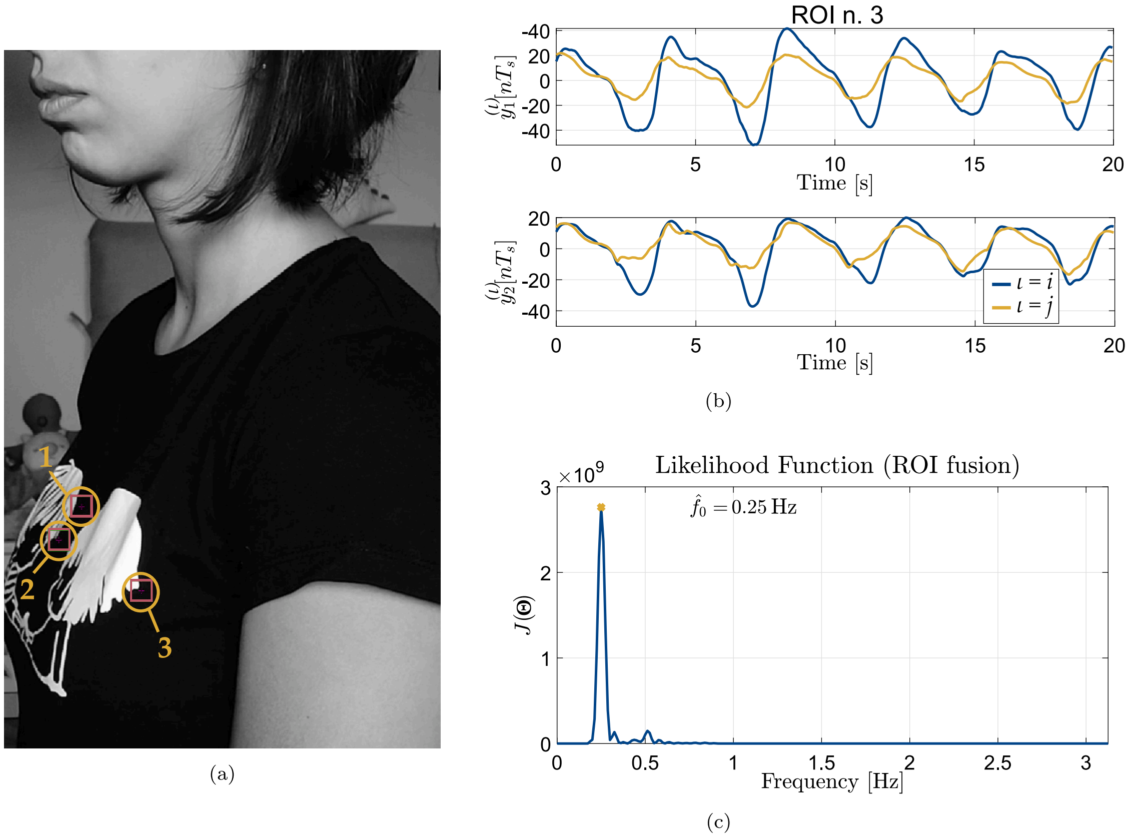
**Algorithm 1** Pseudo-code of the proposed phase-based estimation method.

```
Input: video sequence $f[\mathbf{u},n]$
$\mathcal{R} \leftarrow \emptyset$
$\mathcal{U} \leftarrow \{(u_1,u_2) : \lfloor W/2 \rfloor \le u_1 \le U_1 - 1 - \lfloor W/2 \rfloor,$
        $\lfloor W/2 \rfloor \le u_2 \le U_2 - 1 - \lfloor W/2 \rfloor\}$
$r = 1$
while $r \le R$ do
    $(q_1,q_2) \leftarrow \text{argmax}_{(u_1,u_2)\in\mathcal{U}}\, \tilde{a}_D[\mathbf{u}]$
    $\mathcal{R} \leftarrow \mathcal{R} \cup \{(q_1,q_2)\}$
    $\mathcal{U} \leftarrow \mathcal{U} \setminus \{(u_1,u_2) : q_1 - \lfloor W/2 \rfloor \le u_1 \le q_1 + \lfloor W/2 \rfloor,$
            $q_2 - \lfloor W/2 \rfloor \le u_2 \le q_2 + \lfloor W/2 \rfloor\}$
    $f_r[\mathbf{u},n] \leftarrow f[\mathbf{u},n]$ for $(u_1,u_2) : q_1 - \lfloor W/2 \rfloor \le u_1 \le q_1 + \lfloor W/2 \rfloor,$
            $q_2 - \lfloor W/2 \rfloor \le u_2 \le q_2 + \lfloor W/2 \rfloor$
    $y_m^{(i)}[n] \leftarrow \frac{1}{U_1 U_2} \sum_{u_1=1}^{U_1-1} \sum_{u_2=1}^{U_2-1} \alpha_m f_m^{(i)}[\mathbf{u},n]$
    $y_m^{(j)}[n] \leftarrow \frac{1}{U_1 U_2} \sum_{u_1=1}^{U_1-1} \sum_{u_2=1}^{U_2-1} \alpha_m f_m^{(j)}[\mathbf{u},n]$
    if $\bar{i}_r[n] > \gamma_{\text{th}}$ then
        $\kappa_r \leftarrow 0$
    else
        $\kappa_r \leftarrow 1$
    end if
    $r \leftarrow r + 1$
end while
Output: $\hat{f}_0 \leftarrow \text{argmax}_{f_{\min} \le f \le f_{\max}} \sum_{r=1}^{R} \kappa_r J_r(\boldsymbol{\Theta})$
```

## 4. Applications and results

The performance of the estimation algorithms presented in Sections 2 and 3 is now discussed on the basis of experimental results directly obtained by applying the proposed methods on three sets of videos specifically recorded. In particular, the first set includes 2 videos of a newborn sleeping face-up [12], whereas the second and third sets include 4 and 16 videos, respectively, of adults sitting still. All videos were recorded indoor by placing a camera laterally or in front of a steady subject normally breathing and not affected by respiratory disorders. The camera distance from the subjects, e.g., between 40 and 80 cm, is such that movements of the chest and abdomen related to respiration are clearly visible in the recorded videos and not hidden by the clothes worn by the subjects. Possible random movements of the subject unrelated to respiration do not significantly affect the performance of the estimation algorithms, as the large motion detection algorithm described in Section 3.2 is employed to discard such movements.

Motion signals are initially extracted and compared with reference data. In the case of the newborn, a pneumogram is used as gold standard device to acquire the reference respiratory waveform by placing an elastic belt around the chest of the subject. In the case of adults, two wearable sensors, namely, Shimmer3 by Shimmer Sensing$^{\text{TM}}$ and Equivital EQ02 LifeMonitor by Equivital$^{\text{TM}}$, are used to record, respectively, the reference accelerometric signal and the



(a)

(b)

(c)

**Fig. 5.** Example of: (a) image of a framed subject where 3 ROIs are highlighted, (b) extracted motion information for $r = 3$ and $m = 1, 2$ and (c) likelihood function where the estimated RR is highlighted by the argument of the peak at 0.25 Hz.

respiratory waveform. A comparison between the proposed amplitude- and phase-based methods is also presented. Results in terms of Root Mean Squared Error (RMSE) between estimated and reference data, normalized to the Root Mean Square (RMS) value of the reference data, are finally presented.

In Fig. 5(a), an illustrative image of a framed subject is shown, highlighting three ROIs as square regions. The centres of the ROIs are computed according to the procedure detailed in Section 3.1 for $R = 3$. In Fig. 5(b), the corresponding motion information extracted by the phase-based motion magnification estimation method is shown. In particular, the signals $y_m^{(\iota)}[nT_s]$, $\iota \in \{i, j\}$, obtained by applying (25) at time instants $nT_s$, are plotted over a 20 s time window for the third ROI, i.e., $r = 3$, and for two pyramid levels, i.e., $m = 1, 2$. Finally, in Fig. 5(c), the corresponding likelihood function in (37) is shown as a function of frequency. The estimated frequency $\hat{f} = 0.25$ Hz is the one corresponding to the maximum peak of the function.

As illustrative examples, motion signals extracted by the amplitude- and phase-based motion magnification estimation methods are shown in Fig. 6 and Fig. 7, respectively, along with the corresponding reference signals. In particular, in Fig. 6 the motion signal extracted from a video of a newborn by applying (6) is plotted over a 20 s time window along with the reference signal, i.e., the pneumogram, for the second level ($m = 1$) of the processed pyramid. Considering that one period of the pneumogram corresponds to a complete respiratory cycle, that involves two main movements (inhalation and exhalation), a good correspondence between the two signals can be observed. On the other hand, the average phase variations extracted by two videos, of a newborn and an adult, are plotted over two 20 s time windows and compared with the corresponding pneumogram and accelerometric signals in Figs. 7(a) and 7(b), respectively. In each case, the two pairs of signals exhibit a comparable periodicity, whereas the differences between the two reference signals, in particular the RR, depend on the employed sensors and on the age of the subject.

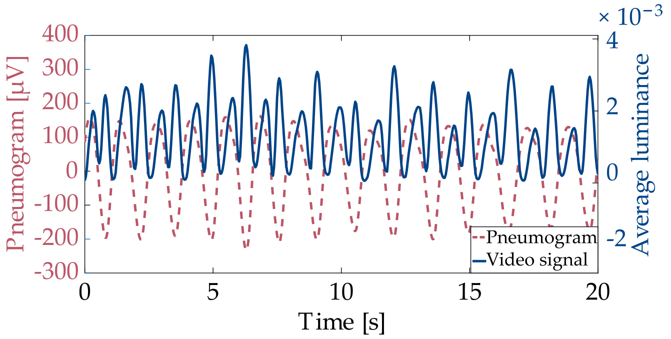
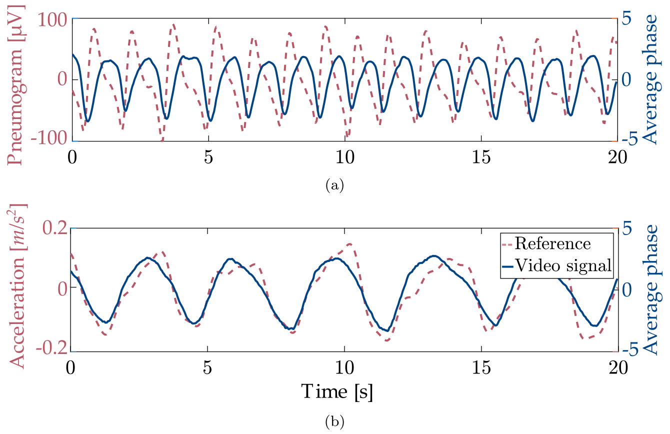


**Fig. 6.** Comparison of the motion signal extracted from a video of a newborn by the amplitude-based motion magnification estimation method and the reference signal, i.e., the pneumogram.

As further investigation, the ML estimation method presented in Section 3 is performed on interlaced windows of $N$ frames, each corresponding to $NT_s$. In the following results, interlaced windows are considered to track the RR over time with proper resolution and the overlap of consecutive windows is defined by an interlacing factor $\rho \in [0, 1)$. An example of windows of length $NT_s$ interlaced by a factor $\rho = 0.75$ is shown in Fig. 8.

In Fig. 9, the frequencies estimated by the phase-based method on interlaced windows for a video framing an adult sitting still are compared with the reference frequencies estimated by the Equivital EQ02 LifeMonitor. The duration of the considered video is 56 s and the RR estimation is performed on 20 s windows interlaced by 90% (i.e., $\rho = 0.9$): this corresponds to 28 processed windows. The first 9 windows should not be considered in the analysis because processed



(a)

(b)

**Fig. 7.** Comparison of two motion signals extracted by the phase-based motion magnification estimation method and the reference signals: (a) pneumogram of a newborn, (b) accelerometric signal of an adult.

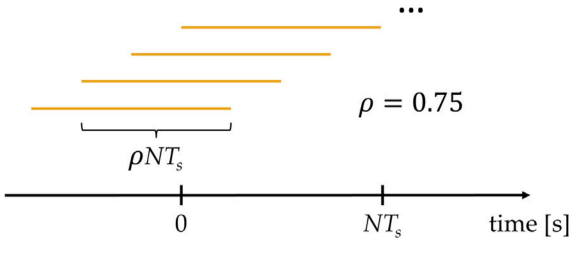


**Fig. 8.** Windows of length $NT_s$ s interlaced by a factor $\rho = 0.75$.

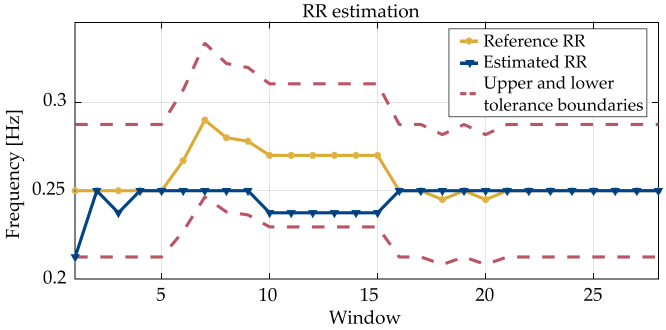
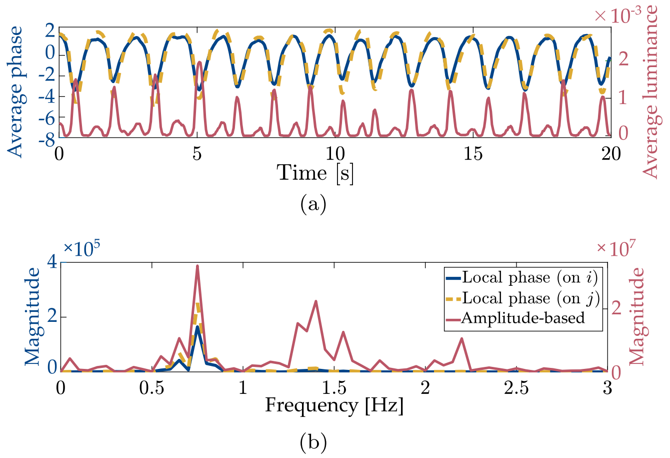


**Fig. 9.** Comparison of estimated and reference RR for the phase-based estimation method.

data are incomplete due to the chosen window overlap pattern. In fact, as shown in Fig. 8 for $\rho = 0.75$, the 3 initial windows are incomplete. It can be noticed that the RR is estimated with good approximation in all windows, confirming the robustness of the system. Tolerance boundaries highlighted in Fig. 9 are computed according to the medical practice of considering acceptable a ±15% variation from to the reference frequency.

A comparison of the presented amplitude- and phase-based methods is now proposed in Fig. 10. In particular, the signal extracted by the amplitude-based motion magnification estimation method from a video of a newborn is shown in Fig. 10(a), along with the corresponding signals $y_0^{(\imath)}[n]$, $\imath \in \{i, j\}$ locally extracted from a selected ROI of the considered video by the phase-based method. The duration of the considered video signal is 20 s. The signal extracted by the amplitude-based method is always positive, as the quantity obtained by applying (6) defines the average luminance for each processed frame. For this reason, inhalation and exhalation acts, which are characterized by movements in opposite directions, may not be clearly distinguishable, especially under critical conditions, e.g., poor camera positioning or patient type. The phase-based method allows to overcome this limitation, as the extracted signals $y_0^{(\imath)}[n]$, $\imath \in \{i, j\}$ exhibit negative values, too.



(a)

(b)

**Fig. 10.** Comparison of the presented methods: (a) extracted motion signals and (b) magnitude spectra.

The different characteristics of the two types of signals are also visible in Fig. 10(b), where their magnitude frequency spectra are plotted. As the phase-based magnification is performed on a selected ROI, the extracted phases are indicated as “local phases” in the legend of Fig. 10(b). A peak around 0.75 Hz can be observed for all the considered cases, corresponding to the correctly estimated RR of 45 breath/min. Nevertheless, the shape of the signal extracted by the amplitude-based method causes other peaks, related to higher order harmonics, to appear around 1.4 Hz and 2.2 Hz. Under critical conditions, these secondary peaks may be higher than the fundamental one impairing RR estimation: for example, a frequency twice the correct one could be estimated. On the other hand, as the signals extracted by the phase-based method are quasi-sinusoidal, due to the direct application of (25), peaks related to higher order harmonics are negligible. This leads to more reliable RR estimation.

### 4.1. Performance analysis

To evaluate the performance of the presented methods, various videos, framing different subjects in different scenarios, are analysed. The main characteristics of the considered videos are summarized in Table 1, where the parameter setting for the video processing analysis is also reported. The durations of the videos vary approximately between 1 min 35 s and 5 min. The camera resolution and the sampling frequency vary according to the employed recording device. We recall that the parameters $M$, $W$, and $R$ indicate, respectively, the number of pyramid levels, the fixed size of the ROIs, and the number of ROIs, the cut-off frequencies of the employed Butterworth filter, used to extract the frequency band of interest, are denoted as $f_{\mathrm{L}}^{\mathrm{co}}$ and $f_{\mathrm{H}}^{\mathrm{co}}$, $\alpha$ is the amplification factor, $NT_s$ is the duration of the processed time window, and the interlacing factor $\rho$ denotes the overlap between consecutive estimation windows. For each video set, the device used as reference is also indicated.

The accuracy of the presented methods is now analysed in terms of normalized RMSE for 6 tested videos (first two sets in Table 1). The results, expressed in dB, are shown in Fig. 11, where the type of framed subject is reported. Considering $N_w$ temporal windows where the RR estimation is performed, the RMSE for each video is defined as

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{n=1}^{N_w} \left|\hat{f}_0[n] - f_0[n]\right|^2}{\sum_{n=1}^{N_w} |f_0[n]|^2}} \qquad (38)$$

**Table 1**
Characteristics of the considered videos and parameter setting.

| Video set | No. videos | Camera resolution | $f_s$ [Hz] | $M$ | $W$ [pixel] | $R$ | $[f_{\mathrm{L}}^{\mathrm{co}}, f_{\mathrm{H}}^{\mathrm{co}}]$ [Hz] | $\alpha$ | $NT_s$ [s] | $\rho$ | Reference device |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Newborns | 2 | 360 × 288 | 25 | 3 | 21 | 4 | [0.3 1.1] | 25 | 20 | 0.5 | Pneumograph |
| Adults | 4 | 800 × 600 | 30 | 4 | 41 | 3 | [0.19 0.9] | 20 | 20 | 0.5 | Accelerometer |
| Adults | 16 | 1920 × 1080 | 30 | 3 | 16 | 3 | [0.19 0.9] | 20 | 20 | 0.5 | Equivital EQ02 LifeMonitor |

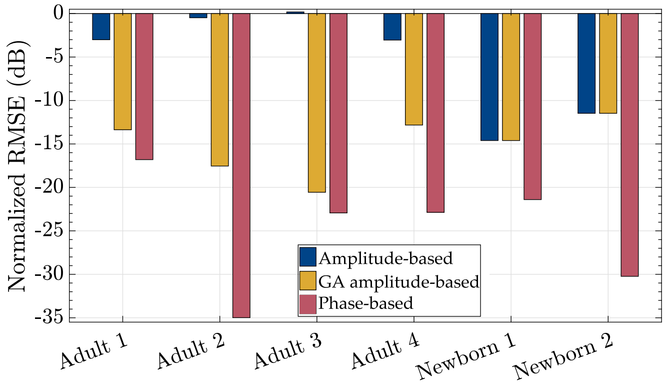


**Fig. 11.** Performance of the assessed methods in terms of normalized RMSE for 6 considered videos (first two sets in Table 1).

where $\{\hat{f}_0[n]\}_{n=1}^{N_w}$ and $\{f_0[n]\}_{n=1}^{N_w}$ are the estimated and reference frequencies for the $n$th window, respectively. The reference frequencies are obtained either by means of an accelerometer (for adults) or a pneumogram (for newborns). As RR estimation based on the amplitude-based approach is prone to errors caused by higher order harmonics, as previously discussed, an idealized Genie-Aided (GA) version of this method is also considered as a benchmark. The GA method automatically corrects estimated double frequencies. Despite this adjustment, the phase-based method exhibits better performance for all the considered videos. Estimates are indeed more reliable due to the characteristics of the motion signals in (25), which inherently allow to distinguish motions in opposite directions associated with inhalation and expiration.

In order to further analyse the performance of the more efficient phase-based method, 16 more videos, all framing adults sitting still, are tested and the normalized RMSE is computed according to (38). Various subjects, scenarios, and camera angles are considered and the Equivital EQ02 LifeMonitor is used as the reference device. The results, expressed in dB, are presented in Fig. 12 and show a good agreement with the RMSE values in Fig. 11, thus confirming the robustness of the considered method. The average error over all the videos is also highlighted as a straight line at −18.7 dB and it can be observed that the RMSE obtained for 9 videos is smaller than or equal to this value.

## 5. Conclusions

In this paper, two contact-less methods to estimate the RR from video sequences are presented. The proposed methods are based on amplitude and phase motion magnification to highlight subtle respiratory movements and combine spatial and temporal processing techniques to extract reliable motion information. Suitable ROIs, where the motion is mainly due to respiration, may be selected to enhance the estimation. Once the motion signals are extracted, the ML principle is applied to estimate, by aggregating data from different ROIs and pyramid levels, the fundamental frequency corresponding to the RR. The accuracy of the two methods is assessed by comparison with reference data, showing good agreement between the estimated signals and the reference ones. Nevertheless, the characteristics of the motion signal obtained by applying the amplitude-based approach may lead to wrong frequency estimates, twice the correct ones. These limitations are overcome by the phase-based method that leads to more reliable estimates due to the regular shape of the extracted motion signals, which may resemble quasi-sinusoidal ones. The performance of the two methods is compared in terms of normalized RMSE showing the higher accuracy of the phase-based approach, which leads to smaller errors for all tested videos.

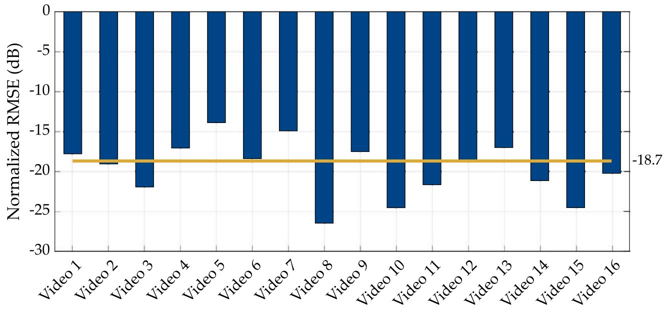


**Fig. 12.** Performance of the phase-based method in terms of normalized RMSE for 16 considered videos (last set in Table 1).

## CRediT authorship contribution statement

**Veronica Mattioli:** Developed the algorithms and the simulation model, Performed the measurements, Ran the computer simulations, Wrote the paper with input from all authors. **Davide Alinovi:** Developed the algorithms and the simulation model, Discussed the results in the initial phase of the project, Wrote the preliminary conference papers. **Gianluigi Ferrari:** Supervised the project from the signal processing viewpoint. **Francesco Pisani:** Supervised the project from the medical viewpoint. **Riccardo Raheli:** Developed the algorithms, Supervised the entire project.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgment

Davide Alinovi received his PhD in 2017 and was a post-doctoral researcher until November 2019, all at the University of Parma. He was a competent, innovative, and passionate researcher who started and carried out the work that led to the preliminary conference papers. Sadly, Davide passed away on September 16, 2020, at the age of 32. His passing left us appalled by the loss of an excellent colleague and dear friend.

## References

[1] I. Costanzo, D. Sen, L. Rhein, U. Guler, Respiratory monitoring: Current state of the art and future roads, IEEE Rev. Biomed. Eng. 15 (2022) 103–121, http://dx.doi.org/10.1109/RBME.2020.3036330.

[2] C. Massaroni, A. Nicolò, M. Sacchetti, E. Schena, Contactless methods for measuring respiratory rate: A review, IEEE Sens. J. 21 (11) (2021) 12821–12839, http://dx.doi.org/10.1109/JSEN.2020.3023486.

[3] C. Massaroni, A. Nicolò, E. Schena, M. Sacchetti, Remote respiratory monitoring in the time of COVID-19, Front. Physiol. 11 (2020) http://dx.doi.org/10.3389/fphys.2020.00635.

[4] J.J. Freundlich, J.C. Erickson, Electrical impedance pneumography for simple nonrestrictive continuous monitoring of respiratory rate, rhythm and tidal volume for surgical patients, Chest 65 (2) (1974) 181–184, http://dx.doi.org/10.1378/chest.65.2.181.

[5] D. Yıldızdaş, H. Yapıcıoğlu, H.L. Yılmaz, Y. Sertdemir, Correlation of simultaneously obtained capillary, venous, and arterial blood gases of patients in a paediatric intensive care unit, Arch. Dis. Child. 89 (2) (2004) 176–180, http://dx.doi.org/10.1136/adc.2002.016261.

[6] B. Schoun, S. Transue, M.-H. Choi, Real-time thermal medium-based breathing analysis with python, in: Proc. 7th Workshop on Python for High-Performance and Scientific Computing, PyHPC'17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 1–9, http://dx.doi.org/10.1145/3149869.3149874.

[7] C.B. Pereira, X. Yu, T. Goos, I. Reiss, T. Orlikowsky, K. Heimann, B. Venema, V. Blazek, S. Leonhardt, D. Teichmann, Noncontact monitoring of respiratory rate in newborn infants using thermal imaging, IEEE Trans. Biomed. Eng. 66 (4) (2019) 1105–1114, http://dx.doi.org/10.1109/TBME.2018.2866878.

[8] C. Massaroni, D.S. Lopes, D.L. Presti, E. Schena, S. Silvestri, Contactless monitoring of breathing patterns and respiratory rate at the pit of the neck: A single camera approach, J. Sens. 2018 (2018) 1–13, http://dx.doi.org/10.1155/2018/4567213.

[9] S. Sanyal, K.K. Nundy, Algorithms for monitoring heart rate and respiratory rate from the video of a user's face, IEEE J. Transl. Eng. Health Med. 6 (2018) 1–11, http://dx.doi.org/10.1109/JTEHM.2018.2818687.

[10] M. Mateu-Mateus, F. Guede-Fernández, N.R.-I. nez, M. García-González, J. Ramos-Castro, M. Fernández-Chimeno, A non-contact camera-based method for respiratory rhythm extraction, Biomed. Signal Process. Control 66 (2021) 102443, http://dx.doi.org/10.1016/j.bspc.2021.102443.

[11] R. Janssen, W. Wang, A. Moço, G. Haan, Video-based respiration monitoring with automatic region of interest detection, Physiol. Meas. 37 (2015) 100–114, http://dx.doi.org/10.1088/0967-3334/37/1/100.

[12] L. Cattani, D. Alinovi, G. Ferrari, R. Raheli, E. Pavlidis, C. Spagnoli, F. Pisani, A wire-free, non-invasive, low-cost video processing-based approach to neonatal apnoea detection, in: 2014 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings, Rome, Italy, 2014, pp. 67–73, http://dx.doi.org/10.1109/BIOMS.2014.6951538.

[13] A. Al-Naji, J. Chahl, Remote respiratory monitoring system based on developing motion magnification technique, Biomed. Signal Process. Control 29 (2016) 1–10, http://dx.doi.org/10.1016/j.bspc.2016.05.002.

[14] S. Alam, S.P.N. Singh, U. Abeyratne, Considerations of handheld respiratory rate estimation via a stabilized video magnification approach, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, Jeju, Korea (South), 2017, pp. 4293–4296, http://dx.doi.org/10.1109/EMBC.2017.8037805.

[15] J. Brieva, H. Ponce, E. Moya-Albor, A contactless respiratory rate estimation method using a Hermite magnification technique and convolutional neural networks, Appl. Sci. 10 (2) (2020) http://dx.doi.org/10.3390/app10020607.

[16] D. Alinovi, G. Ferrari, F. Pisani, R. Raheli, Markov chain modeling and simulation of breathing patterns, Biomed. Signal Process. Control 33 (2017) 245–254, http://dx.doi.org/10.1016/j.bspc.2016.12.002.

[17] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, W. Freeman, Eulerian video magnification for revealing subtle changes in the world, ACM Trans. Graph. 31 (4) (2012) http://dx.doi.org/10.1145/2185520.2185561.

[18] N. Wadhwa, M. Rubinstein, F. Durand, W.T. Freeman, Riesz pyramids for fast phase-based video magnification, in: 2014 IEEE International Conference on Computational Photography, ICCP, Santa Clara, CA, USA, 2014, pp. 1–10, http://dx.doi.org/10.1109/ICCPHOT.2014.6831820.

[19] P. Burt, E. Adelson, The Laplacian pyramid as a compact image code, IEEE Trans. Commun. 31 (4) (1983) 532–540, http://dx.doi.org/10.1109/TCOM.1983.1095851.

[20] S.M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, first ed., Prentice Hall, Upper Saddle River, NJ, USA, 1993.

[21] D. Alinovi, L. Cattani, G. Ferrari, F. Pisani, R. Raheli, Spatio-temporal video processing for respiratory rate estimation, in: 2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings, Turin, Italy, 2015, pp. 12–17, http://dx.doi.org/10.1109/MeMeA.2015.7145164.

[22] D. Alinovi, G. Ferrari, F. Pisani, R. Raheli, Respiratory rate monitoring by maximum likelihood video processing, in: 2016 IEEE International Symposium on Signal Processing and Information Technology, ISSPIT, Limassol, Cyprus, 2016, pp. 172–177, http://dx.doi.org/10.1109/ISSPIT.2016.7886029.

[23] D. Alinovi, G. Ferrari, F. Pisani, R. Raheli, Respiratory rate monitoring by video processing using local motion magnification, in: 2018 26th European Signal Processing Conference, EUSIPCO, Rome, Italy, 2018, pp. 1780–1784, http://dx.doi.org/10.23919/EUSIPCO.2018.8553066.

[24] C. Solomon, T. Breckon, Fundamentals of Digital Image Processing, first ed., Wiley-Blackwell, Croydon, UK, 2011.

[25] A.V. Oppenheim, J.R. Buck, R.W. Schafer, Discrete-Time Signal Processing, third ed., Pearson - Prentice Hall, Croydon, UK, 2010.

[26] N. Wadhwa, H.-Y. Wu, A. Davis, M. Rubinstein, E. Shih, G.J. Mysore, J.G. Chen, O. Buyukozturk, J.V. Guttag, W.T. Freeman, F. Durand, Eulerian video magnification and analysis, Commun. ACM 60 (1) (2016) 87–95, http://dx.doi.org/10.1145/3015573.

[27] M. Felsberg, G. Sommer, The monogenic signal, IEEE Trans. Signal Process. 49 (12) (2001) 3136–3144, http://dx.doi.org/10.1109/78.969520.

[28] M. Unser, D. Sage, D. Van De Ville, Multiresolution monogenic signal analysis using the Riesz–Laplace wavelet transform, IEEE Trans. Image Process. 18 (11) (2009) 2402–2418, http://dx.doi.org/10.1109/TIP.2009.2027628.

[29] N. Wadhwa, M. Rubinstein, F. Durand, W. Freeman, Quaternionic Representation of the Riesz Pyramid for Video Magnification, Tech. rep., MIT, Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA, 2014, URL https://people.csail.mit.edu/mrub/papers/RieszPyr_Quaternion_TechReport.pdf.

[30] N. Wadhwa, M. Rubinstein, F. Durand, W.T. Freeman, Phase-based video motion processing, ACM Trans. Graph. 32 (4) (2013) http://dx.doi.org/10.1145/2461912.2461966.

[31] L. Cattani, D. Alinovi, G. Ferrari, R. Raheli, E. Pavlidis, C. Spagnoli, F. Pisani, Monitoring infants by automatic video processing: A unified approach to motion analysis, Comput. Biol. Med. 80 (2017) 158–165, http://dx.doi.org/10.1016/j.compbiomed.2016.11.010.

[32] N. Patwari, J. Wilson, S. Ananthanarayanan, S.K. Kasera, D.R. Westenskow, Monitoring breathing via signal strength in wireless networks, IEEE Trans. Mob. Comput. 13 (8) (2014) 1774–1786, http://dx.doi.org/10.1109/TMC.2013.117.