

Received March 13, 2022, accepted April 19, 2022, date of publication April 25, 2022, date of current version May 16, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3170052

# A Blind Filtering Framework for Noisy Neonatal Chest Sounds

**DAVOOD FATTAHI**<sup>1,2</sup>, **REZA SAMENI**<sup>3</sup>, (Senior Member, IEEE),  
**ETHAN GROOBY**<sup>1,4</sup>, (Member, IEEE), **KENNETH TAN**<sup>5,6</sup>,  
**LINDSAY ZHOU**<sup>5,6</sup>, **ARRABELLA KING**<sup>5,6</sup>, **ASHWIN RAMANATHAN**<sup>5,6</sup>,  
**ATUL MALHOTRA**<sup>1,5,6</sup>, AND **FAEZEH MARZBANRAD**<sup>1</sup>, (Senior Member, IEEE)

<sup>1</sup>Department of Electrical and Computer Systems Engineering, Monash University, Clayton, VIC 3800, Australia

<sup>2</sup>School of Electrical and Computer Engineering, Shiraz University, Shiraz 71348-51154, Iran

<sup>3</sup>Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA 30322, USA

<sup>4</sup>Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada

<sup>5</sup>Monash Newborn, Monash Children's Hospital, Clayton, VIC 3168, Australia

<sup>6</sup>Department of Paediatrics, Monash University, Clayton, VIC 3168, Australia

Corresponding author: Faezeh Marzbanrad (faezeh.marzbanrad@monash.edu)

This work was supported in part by the Monash Institute of Medical Engineering (MIME) and in part by the Monash Data Future Institute (MDFI). The work of Ethan Grooby was supported by the MIME-Monash Partners-Commonwealth Scientific and Industrial Research Organisation (CSIRO) Sponsored Ph.D. Research Support and Australian Government Research Training Program (RTP) Scholarship. The work of Atul Malhotra was supported by the Royal Australasian College of Physicians Research Fellowship. The work of Faezeh Marzbanrad was supported by the Advancing Women's Research Success Grant Program.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Monash Health Human Research Ethics Committee under Application No. HREA/18/MonH/471.

**ABSTRACT** Chest sound—as the first and most commonly available vital signal for newborns—contains affluent information about their cardiac and respiratory health. However, neonatal lung sound auscultation is currently challenging and often unreliable due to the noise and interference, particularly for preterm infants. The noise often overlaps with the heart and lung contents in both time and frequency. Moreover, the frequency band of the useful components varies from one case to another, making it difficult to separate by fixed band-pass filtering. In this study, a single-channel Blind Source Separation (SCBSS) framework is proposed to separate newborns' lung and heart sounds from noisy chest sounds recorded by a digital stethoscope. This method first decomposes the signal into a multi-resolution representation using a time-frequency transform, and then applies source separation algorithms, to find proper ad hoc frequency filters. In the simulation scenario, two different time-frequency transforms are considered; Stationary Wavelet Transform (SWT) with dyadic bases, and Continuous Wavelet Transform (CWT) with redundant bases. The transforms are followed by three different source separation methods, namely Principal Component Analysis (PCA), Periodic Component Analysis ( $\pi$ CA), and Second Order Blind Identification (SOBI). The yielded combinations are applied to the chest sounds recorded from ninety-one preterm and full-term newborns. The results show that compared to raw signals, fixed band-pass filtering and seven other separation methods, the heart and lung sounds extracted by the proposed methods have higher quality index and also result in more reliable heart and respiratory rate estimation.

**INDEX TERMS** Neonatal chest sound analysis, single-channel blind source separation, heart sound filtering, lung sound filtering.

## I. INTRODUCTION

Diagnosing cardiac and respiratory diseases in newborns have been the focus of many studies in recent years [1]–[3].

The associate editor coordinating the review of this manuscript and approving it for publication was Easter Selvan Suvisheshamuthu.

Timely diagnosis of cardiac issues in newborns tremendously helps physicians to arrange the necessary medical interventions and treatments to prevent irreversible consequences. To assist with diagnosis, there has been a growing interest in using digital stethoscope (DS) and dedicated computer analysis, particularly in pediatrics [4]–[7]. The major advantages

**TABLE 1.** Frequency bands used in heart sound filtering literature.

Authors	Age Group	Freq. Band
El-Segaier et. al. [14], [15]	infants, children	40–300 Hz
Oliveira et. al. [16]	different ages	25–400 Hz
de Vos et. al. [17]	infants, children	20–420 Hz
Emmanouilidou et. al. [18]	infants, children	20–150 Hz
Amiri et. al. [19], [20]	infant	50–200 Hz
Thiyagaraja et. al. [21]	adults	50–250 Hz
Djebbari et. al. [22]	adults	50–300 Hz
Kumar et. al. [23]	adults	1–600 Hz
Ghassemian et. al. [24]	infants	1–200 Hz
Ghaffari et. al. [25]	infants	20–200 Hz

**TABLE 2.** Frequency bands used in lung sound filtering literature.

Authors	Age Group	Freq. Band
Emmanouilidou et. al. [18]	infants , children	50–2500 Hz
K. Shirota et. al. [26]	infants, children	100–700 Hz
Kanga et. al. [27]	infants, children	100–1200 Hz
Pasterkamp et. al. [28]	infants, older ages	100–2000 Hz
T. Imai et.al. [29]	infants	100–700 Hz
Gavriely et. al. [30]	adults	10–1000 Hz
Vannuccini et. al. [31]	different ages	60–2000 Hz
Charbonneau et. al. [32]	adults	50–1200 Hz
Bahoura M. et. al. [33]	adults	60–700 Hz
Reichert et. al. [34]	adults	100–1000 Hz

of DS include its accessibility, affordability, portability and compatibility with smartphones, which make it suitable to be utilized in resource-constrained and undeveloped countries. It also facilitates telemedicine applications for increasing specialist access, clinical decision support, crowdsourcing and home-based monitoring of cardio-respiratory conditions [7]. The feasibility of using DS for neonatal population has been recently investigated in a few studies focusing on heart, lung or abdominal sound auscultation [8]–[13]. However, its application for newborns has been limited in current clinical practice, mainly due to the weakness of the sounds and high levels of interference and background noise [7]. The sounds recorded from newborns often have poor quality and therefore denoising is required prior to further analysis and diagnosis.

Generally, the chest sound is a mixture of heart and lung sounds, as the two clinically remarkable components in chest records, and noises from other internal and external sources. To date, most studies have used typical band-pass filtering to separate these components. However, the selected frequency bands have been inconsistent in different studies, complicating the use of fixed band-pass filtering and the choice of proper cut-off frequencies. Tables 1 and 2 summarize some of the frequency bands used in previous studies, demonstrating their noticeable disagreement. It can be even more challenging for neonatal chest sounds, where the frequency components of lung sounds are affected by the gestational age at birth, lung fluid, and differences in lung, airway and chest wall size and characteristics [2], [3], [12].

One of the potential solutions for separating the components of monophonic records is to use single-channel blind

source separation (SCBSS) methods. These methods usually involve a decomposition procedure followed by a source separation algorithm to find the components of interest [35]. SCBSS methods have been widely utilised in the field of acoustics, including speech recognition [36] and music analysis [37]. They have also been used for biomedical audio signal processing such as fetal-maternal Phonocardiogram (PCG) separation [38].

The most prevalent heart-lung sound separation algorithms in the literature are based on non-negative matrix factorization (NMF) [39], empirical mode decomposition (EMD) [40]–[42], and singular spectral analysis (SSA) [43]. In regard to EMD-based methods, a practical issue is their reconstruction quality. The intrinsic decomposition in EMD does not have any implication on reconstruction optimality and adding optimality constraints requires additional computational cost, which contradict with the simplicity and intuitiveness of EMD [44].

To date, the existing methods in the lung-heart sound literature do not provide a frequency analysis along with their signal separation procedure. Also, to the best of our knowledge, the temporal structures of the sound sources— such as pseudo-periodicity— have not been utilized for the separation of lung and heart sounds. In addition, most of the existing methods have been evaluated only on synthetic data instead of real recordings acquired in real environments. Furthermore, prior research have focused on extracting one of the sources (heart or lung sound), or benefit from synchronous ECG records. While ECG monitoring is standard in newborn intensive care units (NICUs), it is typically not available in all clinical settings, especially in the less affluent regions such as low- and middle-income countries (LMICs).

In this study, SCBSS methods are epitomized and represented in a unified framework. SCBSS are further interpreted as a versatile approach for denoising, filtering and also extracting the heart and lung sounds from noisy neonatal chest audio recordings. Overall, the single-channel signal is decomposed into its frequency sub-bands using a time-frequency decomposition. Then, a linear combination of the frequency channels is found by source separation, such that a measure of decorrelation or independence is maximized. Therefore, the optimal frequency filters are adjusted particularly for each case using SCBSS methods. The framework may be adapted for different applications and conditions by defining suitable discrimination functions and optimizing them. In this study, based on the intrinsic features of the heart and lung sounds, three different separation measures are examined for extracting the sources, including minimum correlation, independent temporal structures, and maximum pseudo-periodicity. However, the methods using different time-frequency decompositions and independence measures are applied on the neonatal chest sound. The results are compared to raw records, fixed band-pass filtered signals, and several state-of-the-art methods. The results are evaluated in terms of an automated signal quality measure, heart-rate and respiratory rate estimation errors, and computational time.

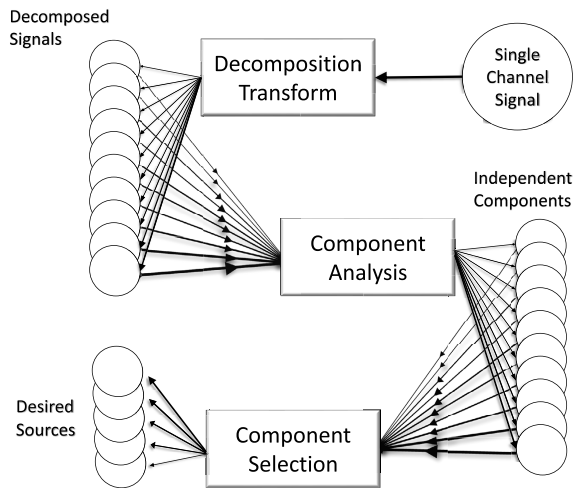


FIGURE 1. The general concept of SCBSS methods.

The following sections are arranged as follows: in Section II, the general concept of tuning the frequency filters for each case and each source is expressed mathematically, and some of its limitations and considerations are discussed. Section III provides details of the chest sounds, the data acquisition procedure, and some possible versions of the proposed framework and their implementation procedures. This is followed by the results of the proposed algorithms in Section IV. Finally, the findings are discussed and concluded in Section V.

## II. BACKGROUND

### A. DATA MODEL AND SCBSS PROBLEM

Suppose that  $x(t) \in \mathbb{R}$  is a single-channel raw chest recording at the time of  $t$ , consisting of three major sources of *heart sound*  $x_H(t)$ , *lung sound*  $x_L(t)$ , and other undesired components and *noise*  $e(t)$ ;

$$x(t) = x_H(t) + x_L(t) + e(t). \quad (1)$$

Separating these sources can be considered as a (blind) source separation problem. However, the solution is feasible if the problem is determined or over determined. It means that the number of observation channels (e.g., simultaneously recording sensors) should be at least equal to or more than the sources. This condition is not met when a single-channel observation exists. Therefore, most SCBSS techniques involve a decomposition process before solving the source separation problem.

Generally, SCBSS methods can be expressed within the framework shown in Fig. 1, where the single channel observation is first decomposed to an  $M$ -channel real or complex signal  $\mathbf{y}(t) \in \mathbb{C}^M$  using a suitable time–frequency transformation, e.g., short-time Fourier transform, wavelet transforms, Hilbert–Huang transforms, etc. [45]. Therefore, each of the decomposed signals in  $\mathbf{y}(t)$  is represented by a convolution of

the observation and the bases:

$$y_j(t) = x(t) * v_j(t) \quad (2)$$

where  $j = 1, \dots, M$  and  $v_j(t)$  denotes the transform basis in the  $j^{\text{th}}$  channel.

Next, source separation methods are applied to separate the independent components. If we consider uncorrelatedness as a weaker measure of independency, we can also use PCA as well as ICA methods (generally abbreviated as “CA” methods in the sequel) to find the sources. Therefore, assuming  $\mathbf{B} \in \mathbb{C}^{M \times M}$  as the matrix of unmixing coefficients, the sources are obtained as:

$$\mathbf{s}(t) = \mathbf{B} \mathbf{y}(t) \quad (3)$$

where  $\mathbf{s}(t) = [s_1(t), \dots, s_M(t)]^T \in \mathbb{C}^M$  denotes the sources vector, while its rows are the linear combinations of interest. Hence,

$$s_i(t) = \mathbf{b}_i^T \mathbf{y}(t) = b_{i1}y_1(t) + \dots + b_{iM}y_M(t), \quad (4)$$

where  $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_M]^T$  and  $b_{ij}$  indicates the elements in  $\mathbf{B}$ . In most applications, one or few of the sources and their corresponding rows in  $\mathbf{B}$  are desired.

For some CA methods the components (sources) are naturally sorted according to a measure, such as energy in PCA or periodicity in  $\pi$ CA. The sorting may help select the desired components, but commonly it is not useful for general ICA methods such as JADE or SOBI. Moreover, the sorting can be unreliable for signals contaminated by harsh non-stationary noise. Therefore, the sources of interests need to be selected generally by a component selection procedure, which can be done empirically or by an automatic algorithm.

The transformed desired sources (denoted by  $\hat{\mathbf{y}}_H$  for heart sound and  $\hat{\mathbf{y}}_L$  for lung sound) can be estimated by mixing the selected components. Suppose that  $\mathbf{A} = \mathbf{B}^\dagger$  is the mixing matrix in CA methods, and  $\hat{\mathbf{s}}_H$  and  $\hat{\mathbf{s}}_L$  are the vector of the desired heart and lung components, with the same size as  $\mathbf{s}$  but in which the unselected components are set to zero. So,

$$\hat{\mathbf{y}}_{H/L}(t) = \mathbf{A} \hat{\mathbf{s}}_{H/L}(t). \quad (5)$$

Then, the inverse decomposition transform of  $\hat{\mathbf{y}}$  can yield the estimated filtered signals in the original domain, i.e.,  $\hat{x}_H(t)$  and  $\hat{x}_L(t)$  for heart and lung sounds, respectively. All the procedure’s steps are summarized in Algorithm 1.

### B. UNCORRELATED AND INDEPENDENT SOURCES SEPARATION

Principle Component Analysis (PCA) and Independent Component Analysis (ICA) are two key techniques to solve source separation problems. For example, PCA seeks to find maximum variance resulting uncorrelated sources. The coefficients vector  $\mathbf{b}_i$  which satisfies the PCA constraint can be estimated by the eigen-vector corresponding to the largest eigenvalues of the observation’s covariance matrix.

However, if being uncorrelated is not a sufficiently strong constraint for separating the sources, ICA methods are the

**Algorithm 1** SCBSS Framework

**REQUIRED:** the single-channel signal, the decomposition transform, the CA method, and the component selection method

**OUTPUT:** the filtered signal,

**PROCEDURE:**

- 1: decompose the 1-d record to an  $M$ -d signal using the decomposition transform,
- 2: analyse the decomposed signals using the CA method and find the independent components,
- 3: for each of the aimed output (heart and lung sounds):
  - using the component selection method, select the desired components and eliminate the rest,
  - reconstruct the decomposed signals by applying the CA inverse transform on the desired components,
  - merge the reconstructed  $M$ -d signals back to the original 1-d domain using the inverse of the decomposition transform.

alternatives. They estimate the sources with higher degrees of independence, but at cost of a higher computational cost, mostly arising from solving a non-linear optimization problem. Most of the ICA methods utilize PCA as a pre-whitening step, and then optimize the ICA contrast function to analyse signal statistics.

Different contrast functions makes different types of ICA methods. For example, JADE algorithm uses the contrast of forth order moments to find the independent sources [46]. Other methods such as AMUSE,  $\pi$ CA and SOBI are satisfied of the second order statistics by diagonalizing the covariance and time-lag auto-covariance matrices of observation [47]–[50]. Therefore, they can be more effective than JADE if the data has temporal structures.

Assume that  $s(t)$  is a zero-mean source signal, and  $E_t\{\cdot\}$  denotes expectation over time, by defining the  $\tau$  time-lagged covariance of  $s(t)$  as

$$C_s(\tau) = E_t\{s(t + \tau)s(t)\}, \tag{6}$$

then it can be expressed using the observation’s covariance matrix:

$$C_y(\tau) = \mathbf{b}^H C_y(\tau) \mathbf{b}, \tag{7}$$

where

$$C_y(\tau) = E_t\{\mathbf{y}(t + \tau)\mathbf{y}(t)^H\}. \tag{8}$$

All the second-order-statistic source separation methods attempt to joint-diagonalize the set of covariance matrices  $\{C_y(\tau_k)\}$  for different time-lag values  $\tau_k \in \{\tau_0(= 0), \tau_1, \tau_2, \dots, \tau_K\}$ , and maximize the following general cost function:

$$\epsilon(\mathbf{b}, \tau) = \sum_{k=0}^K |\mathbf{b}^H C_y(\tau_k) \mathbf{b}|. \tag{9}$$

Therefore, in PCA  $C_y(\tau_k)$  is diagonalized just for  $\tau_0 = 0$ , in AMUSE two covariance matrices with  $k = 0, 1$  are involved, and SOBI utilizes several time-lags i.e.,  $k = 1, 2, 3, \dots, K$ , and attempts to approximately joint-diagonalize their corresponding auto-covariance matrices.

Among the mentioned methods,  $\pi$ CA is the one which by definition is customized for extracting semi-periodic signals. Some of the semi-periodic bio-signals like ECG and PCG can be considered *cyclostationary*, meaning that their statistical characteristics are repeated by a period of time like  $\tau$ . In the wide-sense definition of a zero mean cyclostationary signal, the statistics are reduced to the covariance, and it is expected that

$$C_s(0) = C_s(\lambda \tau), \quad \lambda = 1, 2, \dots \tag{10}$$

This periodicity as a temporal structure can be a useful piece of prior information in coordinating the CA methods to extract the semi-periodic sources. This objective is carried out in  $\pi$ CA by minimizing the following measure of periodicity [48]:

$$\epsilon(\mathbf{b}, \tau) = \frac{\sum_t |s(t + \tau) - s(t)|^2}{\sum_t |s(t)|^2}. \tag{11}$$

It can be shown that the above cost function is equivalent to the following [48]:

$$\epsilon(\mathbf{b}, \tau) = 2 \left[ 1 - \frac{\mathbf{b}^H C_y(\tau) \mathbf{b}}{\mathbf{b}^H C_y(0) \mathbf{b}} \right]. \tag{12}$$

It is obvious that minimizing this measure can be satisfied by the cyclostationarity criteria in (10) for  $\lambda = 1$ . However, this is a Rayleigh–Ritz optimization problem and the value of  $\mathbf{b}$  minimizing (11) and (12) is attainable by the eigenvector corresponding to the largest generalized eigenvalue of the matrix pair  $(C_y(\tau), C_y(0))$ .

**C. FREQUENCY FILTERING BY SCBSS**

Considering (2) and (4), we can rewrite the  $i^{\text{th}}$  source as below:

$$\begin{aligned} s_i(t) &= b_{i1} y_1(t) + \dots + b_{iM} y_M(t) \\ &= b_{i1} [v_1(t) * x(t)] + \dots + b_{iM} [v_M(t) * x(t)] \\ &= [b_{i1} v_1(t) + \dots + b_{iM} v_M(t)] * x(t) \\ &= w_i(t) * x(t) \end{aligned} \tag{13}$$

where  $w_i = \sum_{j=1}^M b_{ij} v_j(t)$  is a linear combination of the decomposition bases. Since the CA methods are linear transformations, if they are applied to time–frequency decomposed signals, the outputs are again frequency filtered signals, but under the constraint of being uncorrelated or (statistically) independent. In other words, they find the coefficients of the frequency filters with filtered components being uncorrelated or independent as much as possible. In the frequency domain, by Fourier transform of (13), we have

$$\begin{aligned} \mathcal{F}\{s_i(t)\} &= [b'_{i1} \mathcal{F}\{v_1(t)\} + \dots + b'_{iM} \mathcal{F}\{v_M(t)\}] \circ \mathcal{F}\{x(t)\} \\ &= \mathcal{F}\{w_i(t)\} \circ \mathcal{F}\{x(t)\} \end{aligned} \tag{14}$$

in which  $\circ$  indicates an element-wise multiplication. So, it is equivalent to frequency filtering of  $x(t)$  using the filter obtained from a linear combination of the time–frequency transform bases. In other word, the coefficients calculated from the CA methods amplify the frequency bins which discriminate the sources, and attenuate the other correlated frequency bins.

#### D. APPROPRIATE DECOMPOSITION TRANSFORMS

For the hereby proposed framework, various choices of time–frequency expansions are possible. Some of them such as Discrete Wavelet Transform (DWT) benefit from orthonormal bases, providing no redundant components. However, some of the DWT specifications should be modified and matched to the proposed SCBSS framework. Firstly, the CA methods need the decomposed signals at each scale to have the same length, which is not valid in the orthonormal expansions like DWT because of its decimated translation. It can be addressed using undecimated versions of DWT like Stationary Wavelet Transform (SWT) [51] or Maximal Overlap Discrete Wavelet Transform (MODWT) [52].

Although the undecimated translation causes the bases in the same scale to be redundant and not orthogonal, they remain mutually orthonormal in different scales. Therefore, the decomposed signals are supposed to be uncorrelated in different levels, and no more de-correlation process can be done by the methods like PCA. For illustration, consider an orthonormal time–frequency transform followed by PCA. Regarding the equation (13), the desired linear combination is estimated by PCA by finding the eigen-vector corresponding to the largest eigenvalue of  $C_y$ , the covariance matrix of  $y(t)$ . However, if orthogonal bases are used in the time–frequency decomposition,  $C_y$  becomes the identity matrix, i.e., PCA does not add any advantage beyond the sub-band decomposition readily accomplished by DWT. However, other CA methods using more potent independence constraints like ICA techniques can still discriminate the sources. Moreover, orthonormal expansion sets involve dyadic scaling, predefined frequency bands and constant Q-factor wavelet decomposition. These restrictions lead to less flexibility in tuning the frequency-domain filtering. To avoid these constraints and to achieve more degrees of freedom, one may employ more redundant transformations, such as the DWT with Biorthonormal bases, Rational Dilation Wavelet Transform (RADWT), or even more redundant expansions such as the Continuous Wavelet Transform (CWT), which benefits from highly localized time–frequency filters. Even more redundancy (viz. degrees of freedom) can be achieved by CWT with complex coefficients, which comprises phase information as well as amplitude. Note that these excess redundancies will be subsequently reduced by the CA methods, and do not imply any restriction but computational cost. The advantages of utilizing redundant transforms have been studied in the context of heart and lung sound processing [53]–[56].

### III. METHODS

#### A. DATA ACQUISITION

In this work, the proposed methods were applied to single-channel chest sounds recorded from two groups of pre-term and full-term infants. To record the sounds, the digital stethoscope (Clinicloud Stethoscope, Clinicloud Ltd Pty, Melbourne, Australia) was placed on the right anterior chest of the infant. Lung and heart sounds were then recorded for 60 s. The recordings were saved by Voice Recorder & Audio Editor [57], a commercially available smartphone software, in MP3 format with 16 kHz or 44 kHz sampling rates. The study was conducted at Monash Newborn, Monash Children Hospital, a tertiary-level neonatal unit in Melbourne, Australia, and was approved by the Monash Health Human Research Ethics Committee (HREA/18/MonH/471). Recordings were made between 24 and 48 hours after birth, to avoid potential interference from lung fluid clearance in the first 24 hours. More information about the data is available at [12], [58].

Some of the recordings were excluded manually because of significant artifacts such as crying, recorder saturation, stethoscope slipping, and some loud environment noises, making the lung and heart sounds impossible to recover. Since irregular breathing periods existed in most of the studied newborns, specially for the preterm cases, there are long non-breathing segments in the recordings. Therefore, 10 s segments containing both the heart and lung sounds were chosen manually by an expert. Totally, 91 segments are considered for further analysis, while 28 of them were accompanied by NICU Heart Rate (HR) and Respiration Rate (RR) information. In addition, to avoid unnecessary high computational cost in future processes, the recordings were down-sampled to 4 kHz.

#### B. ALGORITHM DEVELOPMENT AND IMPLEMENTATION

Different choices for the decomposition transform and the CA method in the proposed framework of Fig. 1 provides a variety of SCBSS algorithms. In the present study, SWT and CWT in combination with PCA, SOBI and  $\pi$ CA methods provide four algorithms to separate lung and heart sounds from single channel chest records. Their configurations and implementation details are described in the following.<sup>1</sup>

##### 1) SWT–PCA

As the first and simplest configuration, we adapted the Biorthogonal basis function with one and three vanishing moments for decomposition and reconstruction wavelets [59], [60]. Moreover, in order to keep the length of the signals unchanged, an undecimated version of DWT known as Stationary Wavelet Transform (SWT) is utilised [51]. Considering the sampling frequency of the data (4 kHz), a 6-level SWT is used to decompose the signal into dyadic subbands shown in Fig. 2. The six detail levels

<sup>1</sup>The implemented codes of the methods are available at: <https://bitbucket.org/fmarzbanrad/blind-filtering-framework>

are given to PCA and the heart and lung sounds are eventually selected among the components. The final desired heart and lung sounds can be achieved according to Step 3 in Algorithm 1.

### 2) CWT-PCA

Discrete Time Continuous Wavelet Transform (DT-CWT) is a redundant time–frequency expansion utilized as the decomposition transform in our framework. Using one sample translation of the bases leads the decomposed signals to have the same length as the original signal. On the other hand, since the bases are redundant in scales, we can skip some of the scales considering a trade-off between the efficient redundancy and the computational cost. In the present study, we use Analytical Morlet (Gabor) wavelets with 17 scales (four voices per octave) covering 50–1000 Hz. Then like the previous algorithm, we can expect that PCA ranks the possible sources corresponding to heart sound, lung sound and noise. Therefore, according to our visual inspection, the best components corresponding to the heart and lung sources are selected, and the desired sources are obtained after reconstructing and converting the coefficients into the original domain.

### 3) CWT-SOBI

SOBI utilizes plenty of time–lags and attempts to joint-diagonalize the corresponding covariance matrices. The covariance matrix of the time–lag  $\tau_k$  is calculated as:

$$\tilde{C}_y(\tau_k) = [\sum_t y(t + \tau_k)y(t)^H]/L, \quad k = 1, \dots, K. \quad (15)$$

In combination with CWT, SOBI is supposed to extract the sources by finding linear combinations of the sub-bands which maximize their temporal dependency. According to a regular use of SOBI, the set of time–lags  $\tau = \{\tau_1, \dots, \tau_K\}$  might be considered randomly or uniformly just to sweep the time interval containing the expected temporal structures. Instead, in the present study, we propose a more proper method to set the time–lags based on the available prior information of heart sounds.

First of all, the space between the time–lags can be inferred from the expected higher bound of heart sounds frequency components, denoted by  $f_h$ . Defining  $\delta\tau = \tau_k - \tau_{k-1}$ , we have  $\delta\tau = \frac{1}{2f_h}$ . Although the maximum frequency is reported about 300 Hz for neonates, a slightly higher value can be chosen to ensure that all the frequency components are covered [14]. Therefore, taking  $f_h \simeq 330$  Hz, the time–lags are chosen uniformly with 1.5 ms spacing.

In addition, it is expected that the autocorrelation of cyclostationary signals have a higher envelope around the shifts equal to the recurring period of the heart sound. Therefore, if the average period of heart sound is  $T$ , the shifts around  $\lambda T$  for  $\lambda = 0, 1, 2, \dots$  are appropriate candidates to be used as SOBI’s time–lags. Mathematically, let  $\tau_{p_\lambda}$  be the time around  $\lambda T$  at which the autocorrelation envelope has the local

maximum, i.e.,

$$\tau_{p_\lambda} = \arg \max_{\tau} \{|R_x(\tau) + \mathcal{H}\{R_x(\tau)\}|\} \quad \text{s.t. } |\tau - \lambda T| \leq r, \quad \lambda = 0, 1, \dots, \Lambda, \quad (16)$$

where  $R_x$  is the autocorrelation of  $x$ ,  $\mathcal{H}\{\cdot\}$  denotes Hilbert transformation, and  $r$  is the neighbourhood radius around the  $\lambda T$  in which we search for the local maximum. So, from a wide set of available  $\tau_k$ , the proper time–lags  $\tau_{k'}$  can be found as

$$\tau_{k'} = \{\tau_k \mid |\tau_{p_\lambda} - \tau_k| \leq \Delta\tau\}, \quad (17)$$

where  $\Delta\tau$  is named maximum effective lag and can be calculated as  $\Delta\tau = \frac{1}{2f_l}$ , where  $f_l$  is the expected lower frequency bound of the desired signal, here the heart sound. By assuming  $f_l = 50$  Hz according to the literature in [19]–[22], the proper time–lags are supposed to be around the peaks of the autocorrelation envelope with at most  $\Delta\tau = 10$  ms difference.

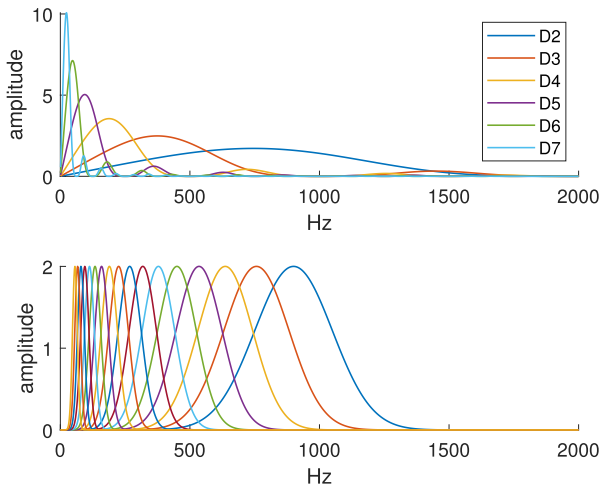
Number of detectable peaks in envelope of auto-correlation is affected by quality and regularity of heart sound. In the present study, the time–lags are set to cover three heart beats, which means  $\Lambda = 3$  in (16).

The CWT configuration is the same as the previous algorithm. After applying the SOBI, the suitable components corresponding to the heart and lung sounds are visually selected from the delivered components, and the rest is performed according to the step 3 of the general framework in Algorithm 1.

### 4) CWT- $\pi$ CA

As mentioned before,  $\pi$ CA unlike SOBI uses a specific time–lag obtained from prior information [48]. In the case of heart sound filtering, it can be related to an averaged or synchronized heart beat captured from an auxiliary signal, like ECG or HRV. However, this reference time–lag is achieved in this study from the chest sound itself without using extra hardware and recording auxiliary signal. The heart sound is the major component in the lower frequencies, and the other components are almost negligible compared with it. So, the signals decomposed by one of the CWT scales corresponding to low frequencies (75–90 Hz) is selected, and its envelope is calculated. Since the Analytical Morlet provides complex coefficients, the envelope can be yielded easily by their absolute values. Then, a conventional peak detection algorithm can find S1 peaks, and the S1–S1 intervals are used as the reference time–lag for  $\pi$ CA. The peak detection algorithm searches for each peak in the expected interval of  $[0.7 T, 1.3 T]$  after the previous peak, while  $T$  is the average heart sound period.

Because of variability in heart rate, the S1-S1 intervals are not equal, and it can decrease the accuracy of calculating  $C_y(\tau)$ . Therefore, in order to make the beats more aligned, we use a simple linear *phase warping* and map each beat to  $(-\pi, \pi)$ . So, the covariance matrices in (12) are calculated



**FIGURE 2.** The frequency sub-bands corresponding to (a) Stationary Wavelet Transform with 6 levels decomposition, (b) Continuous Wavelet Transform with 4 voices per octave.

in the phase domain as:

$$\tilde{\mathbf{C}}_y(\tau) = \left( \sum_t \mathbf{y}(t + \tau) \mathbf{y}(t)^H \right) / L \quad (18)$$

in which  $L$  is the total number of samples, and

$$\tau_t = \min\{\tau | \phi(t + \tau) = \phi(t), \tau > 0\} \quad (19)$$

where  $\phi(t)$  is the *cardiac phase* signal [48]. To ensure that the eigenvalues are real, we can make the covariance matrix symmetric by

$$\tilde{\mathbf{C}}_y(\tau) = (\tilde{\mathbf{C}}_y(\tau) + \tilde{\mathbf{C}}_y(\tau)^H) / 2. \quad (20)$$

By descending sort of the generalized eigenvalues of the pencil  $(\tilde{\mathbf{C}}_y(\tau), \tilde{\mathbf{C}}_y(0))$  the sources are ranked from high to low periodicity. Since the heart sound is the most periodic source, we can expect to see it among the first components, and the lung sound is found in other components. However, because of some occasional non-stationary noises in the recorded data, this order may be slightly inconsistent.

### C. COMPONENT SELECTION

The component selection can be carried out by classifying the components into three group of *heart sound*, *lung sound* or *noise*. The following features are used as inputs:

- peak frequency of power spectral density,
- band power ratio of 50–200 Hz (as the heart sound dominant frequency band) to 50–2000 Hz,
- band power ratio of 200–400 Hz (as the heart and lung sounds mixture frequency band) to 50–2000 Hz,
- band power ratio of 400–1000 Hz (as the lung sound dominant frequency band) to 50–2000 Hz,
- band power ratio of 1000–1200 Hz (as the lung sound and noise mixture frequency band) to 50–2000 Hz,
- band power ratio of 1200–2000 Hz (as the noise dominant frequency band) to 50–2000 Hz,

- standard deviation of S1-S1 intervals, as a periodicity related feature. The S1 detection algorithm is same as the one used in CWT- $\pi$ CA algorithm.
- ratio of the second peak to the first peak of the envelope of auto-correlation absolute values. As a periodicity related feature, this value falls between 0 (least periodicity) and 1 (most periodicity). The procedure of finding first and second peaks is similar to (16), while  $\lambda = 0, 1$ .

15 chest records were utilised here to optimize the classifier, and then excluded from the main dataset in the subsequent performance evaluation. So, 840 components were produced using the 4 proposed algorithms and labeled via listening and visual assessing by an expert, to be used as the training dataset. Five popular classifiers including K-Nearest Neighbours (KNN), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Naive Bayesian, and multi-class Binary Decision Tree (BDT) were applied and their parameters were optimized in a 10-fold cross validation. Then, their classification accuracy rates were obtained through a 10-by-10 cross validation, and KNN was chosen according to its better result (see Table 3).

The component selector may occasionally label none of the components as one or two of the desired sources. In the present study, it happened for heart sound extraction of 3.4 % of the recordings, but none of the lung sound extractions. However, in the case of no-labeling, in order to avoid empty outputs, the components with the power peak closest to 50 Hz and 400 Hz are returned respectively as heart and lung sounds.

### D. PERFORMANCE EVALUATION

#### 1) COMPARED METHODS

To assess the improvement of heart and lung sound quality by the proposed methods, they are compared to the raw sounds and the fixed band-pass filtered signals. The band-pass filters are 10<sup>th</sup> order Butterworth passing 100–1000 Hz for the lung sounds and 50–250 Hz for the heart sounds. The cutoffs are corresponding to the dominant frequency bands of each signal based on the previous studies listed in Tables 1 and 2. In addition, seven state-of-the-art heart-lung separation methods are examined and compared to the proposed methods including: Singular Spectrum Analysis (SSA) [43], Empirical Mode Decomposition (EMD) [40], Ensemble Empirical Mode Decomposition (EEMD) [41], Complete Ensemble Empirical Mode Decomposition (CEEMD) [42], Non-negative Matrix Factorization Clustering with Kullback-Leibler (KL) divergence cost function (NMFC-KL), Non-negative Matrix Factorization Clustering with L-2 cost function (NMFC-L2), and Non-negative Matrix Factorization Clustering with Sparse cost function (NMFC-Sparse) [39], [61], [62]. The parameters and settings of these methods are chosen based on the suggested values in the corresponding references. The utilised toolboxes and codes can be found in [62]–[66].

**TABLE 3.** Accuracy rate of component selectors, shown by mean  $\pm$  standard deviation of accuracy percentage in 10-by-10 cross validation.

Classifier	KNN	LDA	SVM	Bayesian	BDT
Accuracy (%)	92.93 $\pm$ 3.25	73.90 $\pm$ 4.33	92.04 $\pm$ 3.03	90.80 $\pm$ 3.61	86.40 $\pm$ 3.66

## 2) SIGNAL QUALITY ASSESSMENT

The measure of quality utilised here for both the heart and lung sounds assessment is based on our previous work in [67]. Briefly, a linear regression model has been tuned in order to map a collection of audio features into a 5-point scale measuring signal quality for heart and lung sounds (from 1 for the lowest to 5 for the highest quality). The quality scores are obtained after annotation of 176 chest sound recordings (different from ones involved in the present study) by 7 experts and used as the ground truth. Different varieties of features are engaged as the predictors, including: several envelope-based and wavelet-based features, statistical characteristics of the signals like variance and kurtosis, mel-frequency cepstrum, linear predictive coding, and line spectral frequencies coefficients, power ratios of different frequency ranges, dominant frequencies, properties of different fractal dimensions and entropy measures, number of detectable heart sounds (S1 and S2) and breathing periods, and length of periods without detectable heart and lung sounds. The best feature set is selected using Minimum Redundancy Maximum Relevance (MRMR) algorithm. Then, the regression model is trained by a set of best features, while different cost functions (SVM and least-squares) and different regularization methods (lasso and rigid) are examined to obtain the best test results based on mean squared error. More description of the features and the regression can be found in [67].

## 3) HR AND RR DETECTION

Estimating the vital signs like HR and RR is expected to be more reliable and accurate in well-filtered chest sounds. Therefore, two HR and BR estimation algorithms are applied on the filtered heart and lung sounds respectively, in addition to the raw signals. The results are then compared to the reference values recorded simultaneously via an NICU machine. The HR estimation algorithm introduced by Springer et. al., is initialized with HR determined from peak detection of envelope, and then uses it for heart segmentation to obtain overall heart rate estimate [68]. Breathing rate is determined from peak detection of 300-450Hz power envelope. More details of calculated heart and breathing rate are described in [58]. Their Root Mean Square Error (RMSE) compared to the NICU reference values are shown in Table 5.

## 4) COMPUTATIONAL TIME

The algorithms are implemented in MATLAB 2020b and Microsoft Windows 10  $\times$  64, on a hardware with Intel<sup>®</sup> Core<sup>™</sup> i5-8250U CPU @ 1.60 GHz, 1800 MHz, 4 Cores, 8 logical processors, and 8.00 GB of installed physical memory (RAM). The required time of extracting both heart and

lung sounds from 10 s recordings are calculated for the proposed and benchmark methods, and their mean and standard deviation are reported in Table 4. For the KNN component selector utilised in the proposed methods, mean and standard deviation of the running time was  $66 \pm 11$  ms, which is included in the table values.

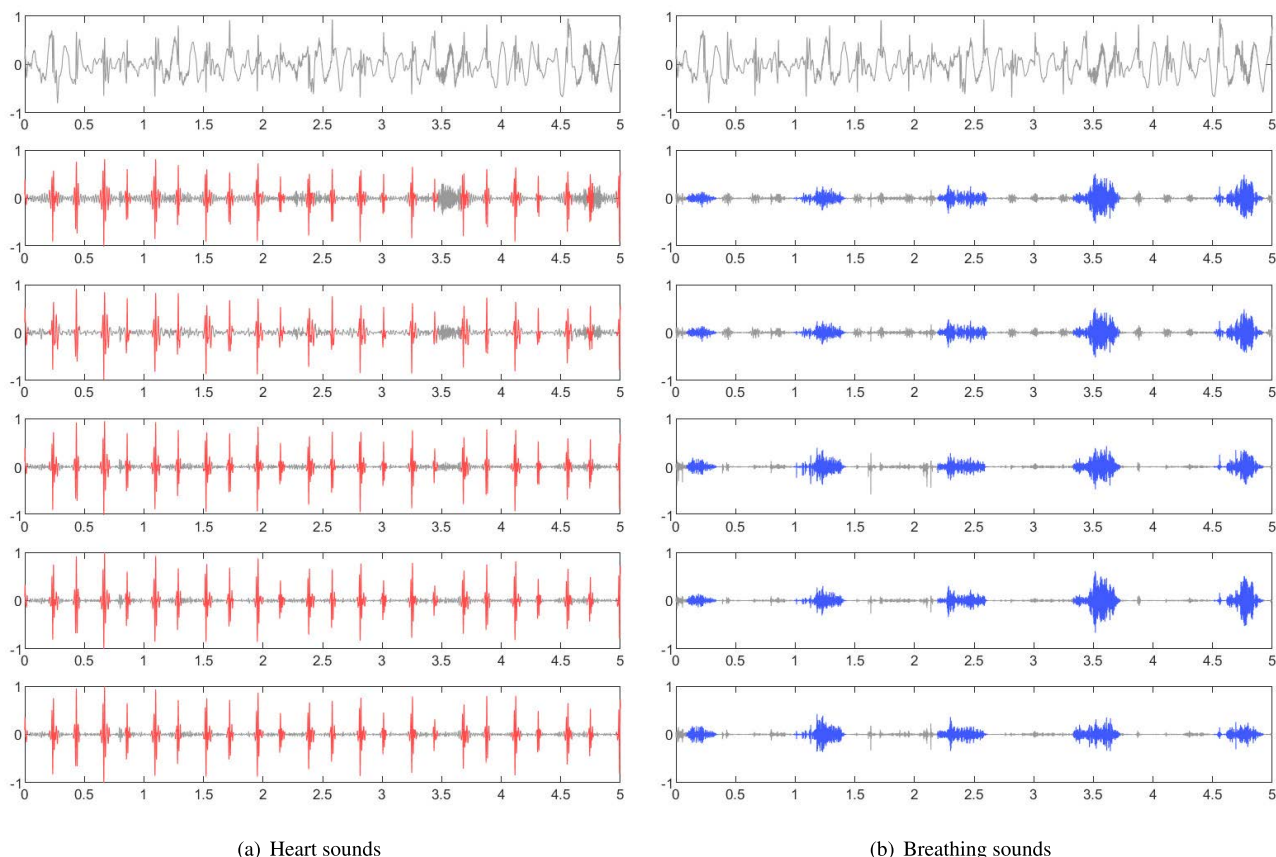
## IV. RESULTS AND DISCUSSION

Fig. 3 shows a 10 s sample chest record which are filtered by 4 different methods proposed in the present study and also band-pass filtering with fixed cut-offs. First of all, we can see that the heart and lung sounds obtained by the fixed band-pass filters are not filtered well, and both of them are contaminated with other undesired components. Clearly, the residual heart sound exists in the filtered lung sound, and vice versa. It to some extent happened for the heart and lung sounds filtered by SWT-PCA, but the amount of undesired components are less than fixed band-pass filtering. However, these residuals are pretty much reduced in the sources estimated by CWT-PCA, CWT- $\pi$ CA and CWT-SOBI.

The quality of the sources extracted by the proposed and compared methods is assessed by the methods described in III-D. Fig. 4 shows how much improvement is caused by each method compared to the raw signals for heart and lung sounds. Obviously, all the proposed methods have provided more quality improvement for the heart and lung sounds compared to the simple band-pass filtering. The performance of CWT-PCA is better than SWT-PCA in lung sound extraction, that may imply the advantage of using redundant bases instead of dyadic ones in high frequencies. The results of CWT- $\pi$ CA were expected to be better than CWT-PCA, but they are not;  $\pi$ CA finds the proper components by diagonalizing just one delayed auto-covariance matrix, and its result is highly sensitive to its time-lag value. Miss-detection of S1 in poor quality sounds or presence of significant non-stationary noises can drastically affect estimating accurate values for period of heart sounds and consequently the time-lag. This drawback would likely be addressed if an external heart beat detection source were used. Using auxiliary signals like ECG along with chest sound could be helpful in this regard. Lately, simultaneous ECG-PCG recording and developing related devices, which have recently become an interesting trend [69], [70]. However, CWT-SOBI are more robust against the mentioned noises, because of using plenty time-lags and joint diagonalization of more than two auto-covariance matrices. Therefore, it can be seen that the results of CWT-SOBI for heart sound quality improvements is much better than others.

Among the other compared algorithms, SSA outperforms the EMD-based and NMF-based methods. Its heart sound





**FIGURE 3.** The sources extracted by different methods from the same 5 s sample of chest sound; (a) filtered heart sounds, the foregrounds in red indicate the annotated S1 and S2 sounds, (b) filtered lung sounds, the foreground in blue is the annotated exhale intervals. From top to bottom: Raw signal, SWT-PCA, CWT-PCA, CWT- $\pi$ CA, CWT-SOBI.

**TABLE 4.** Computational time of extracting heart and lung sounds from a 10 s chest record using different methods. The component selection time is included for the proposed methods.

Method	SWT-PCA	CWT-PCA	CWT- $\pi$ CA	CWT-SOBI	SSA	EMD	EEMD	CEEMD	NMFC-KL	NMFC-L2	NMFC-Sparse
time(ms)	256 ± 18	473 ± 19	498 ± 29	2211 ± 881	694 ± 39	1170 ± 130	5948 ± 569	10481 ± 1041	1147 ± 33	510 ± 21	1128 ± 29

SQI is almost like CWT-SOBI, but its lung sound SQI is not competitive with the proposed methods. EMD-based and NMF-based methods are in the next ranks for both heart and lung sound SQIs. It is also observed that there are occasional degradations in all the proposed and benchmark methods. This can be explained by considering that: A) the utilized SQI estimation method is not 100 % accurate and occasional incorrect estimations would be expected. But on average it is an appropriate metric for SQI calculation; B) None of the proposed and benchmark algorithms are perfect in their performance, especially for the case of low quality and noisy chest sound of newborns.

Table 5 shows the effect of different filtering methods on accuracy of HR and RR estimation. As we can see, CWT- $\pi$ CA, CWT-PCA, and CWT-SOBI generally caused lower RMSE for HR and BR estimation than the benchmark

methods, fixed band-pass filtering, and raw signals. However, no improvements in RR and HR is demonstrated by SWT-PCA. It might be caused by its dyadic sub-bands which leads less flexibility of filter adjustment specially in low frequencies.

The computational time of the proposed methods for a heart-lung sound extraction from a 10s record are demonstrated in Table 4. CWT-SOBI has significantly higher needed time compared to the other proposed methods, because of its joint diagonalization of plenty covariance matrices. Even so, it seems feasible to use the algorithms in commercial, portable and affordable medical devices. Among the seven benchmark methods, SSA and NMFC-L2 also have relatively low computational time, while the worst time-consuming method is CEEMD.

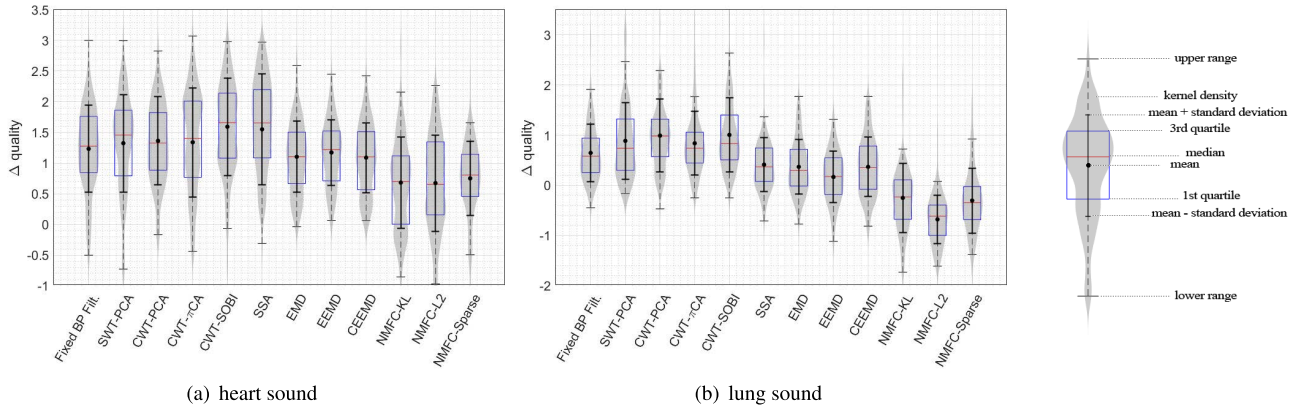


FIGURE 4. Signal quality index (SQI) improvement of heart and lung sounds by employing different algorithms vs the raw records.

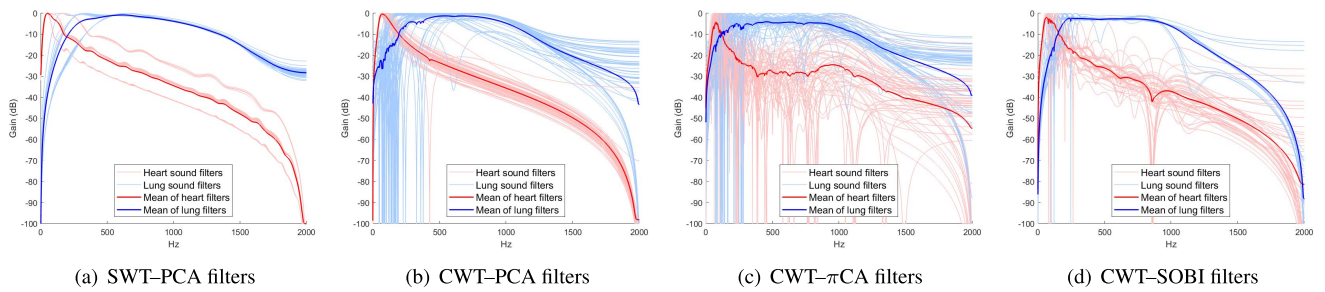


FIGURE 5. The filters optimized by the proposed methods, depicted in logarithmic scales; blue for lung sound filtering, and red for heart sound filtering. The overlaid transparent lines represent the filters for different cases, and the bold one represents their mean value.

TABLE 5. RMSE of the estimated HR (beats per 10 s) and RR (count per 10 s) of the extracted heart and lung sounds, and raw signals, compared to the NICU reference values.

Method	Raw	Fixed BP Filt.	SWT-PCA	CWT-PCA	CWT- $\pi$ CA	CWT-SOBI	SSA	EMD	EEMD	CEEMD	NMFC-KL	NMFC-L2	NMFC-Sparse
HR error	1.68	1.51	2.02	0.88	<b>0.65</b>	0.79	1.68	1.42	1.82	1.39	1.62	2.00	1.62
RR error	2.43	2.30	2.40	2.27	<b>2.17</b>	2.19	2.25	2.43	2.53	2.43	2.24	2.50	2.44

Fig. 5 depicts the frequency filters adjusted for extracting heart and lung sounds using the four implemented methods. The filters of all the records are overlaid on each other in order to show their variation caused by the adaptability of the methods. The SWT-PCA filters has the least variation because of less frequency localization of SWT compared with CWT. The  $\pi$ CA filters seems unstable and tumultuous, which as mentioned before, might be caused by the inappropriate setting of the time-lag.

V. CONCLUSION

In auscultation of chest recordings, usually either the heart or lung sounds are the intended acoustic signals for examination. However, the recording inevitably is a mixture of both components and also different environmental and acquisition noises, causing lower agreement and accuracy in medical diagnosis. Therefore, filtering of the raw signals is a crucial step before any clinical auscultation. We have found it particularly challenging for newborns’ chest auscultation, due to

the closeness of the heart and the lung, noise and weakness of the sounds of interest.

Despite the common use of frequency filtering, characteristics of an “optimal” frequency filter varies among the cases. Specifically, defining suitable cut-off frequencies has not been straight forward. We addressed this challenge by customizing a SCBSS framework using the combination of different CA methods and time-frequency expansions to design ad-hoc filters, per case. Its main advantage is that it can find the adaptable filters for each case in a non-parameter procedure. In addition, using the proposed framework both heart and lung sounds can be extracted simultaneously. In two of the methods suggested within the framework, periodicity and temporal structure of the heart sound is used as the contrast function in tuning the frequency filters. Specifically, a novel method is proposed to set the parameters of SOBI (specifically the proper time-lags of its lagged covariance matrices) according to known characteristics of heart sound such as its lower and higher frequency bounds. The results of

implementation using neonatal chest sounds showed that the proposed methods can improve the quality of both heart and lung sounds, compared to seven state-of-the-art heart-lung sound separation algorithms, fixed bandpass filtering and the raw signals.

The proposed methods can potentially assist healthcare providers with a more reliable auscultation using an inexpensive DS, applicable also in resource-limited settings. It is more significant in the potential application in tele-health, when the sounds are recorded by non-experts, for remote monitoring and diagnosis.

The present study will be more comprehensive and enriched by setting out the following point for further study; Involving other decomposition and CA methods may provide higher source extraction performance. Assaying the methods on other older age groups, including children and adults, and also the cases with pathological heart and lung sounds, potentially from the same group of subjects, could be investigated in future studies.

## REFERENCES

- [1] A. M. Rudolph, "Diagnosis and treatment: Respiratory distress and cardiac disease in infancy," *Pediatrics*, vol. 35, no. 6, pp. 999–1002, 1965.
- [2] J. B. Warren and J. M. Anderson, "Newborn respiratory disorders," *Pediatrics Rev.*, vol. 31, no. 12, pp. 487–495, Dec. 2010.
- [3] H. Aly, "Respiratory disorders in the newborn: Identification and diagnosis," *Pediatrics Rev.*, vol. 25, no. 6, pp. 201–208, Jun. 2004.
- [4] E.-M. Strehle, G. Earle, B. Bateman, and K. Dickinson, "Teaching medical students pediatric cardiovascular examination by telemedicine," *Telemedicine e-Health*, vol. 15, no. 4, pp. 342–346, May 2009.
- [5] A. Garingo, P. Friedlich, L. Tesoriero, S. Patil, P. Jackson, and I. Seri, "The use of mobile robotic telemedicine technology in the neonatal intensive care unit," *J. Perinatol.*, vol. 32, no. 1, pp. 55–63, Jan. 2012.
- [6] L. E. Ellington, R. H. Gilman, J. M. Tielsch, M. Steinhoff, D. Figueroa, S. Rodriguez, B. Caffo, B. Tracey, M. Elhilali, J. West, and W. Checkley, "Computerised lung sound analysis to improve the specificity of paediatric pneumonia diagnosis in resource-poor settings: Protocol and methods for an observational study," *BMJ Open*, vol. 2, no. 1, 2012, Art. no. e000506.
- [7] A. Ramanathan, L. Zhou, F. Marzbanrad, R. Roseby, K. Tan, A. Kevat, and A. Malhotra, "Digital stethoscopes in paediatric medicine," *Acta Paediatrica*, vol. 108, no. 5, pp. 814–822, 2019.
- [8] B. P. Treston, J. Semberova, R. Kernan, E. Crothers, A. Branagan, N. O' Cathain, and J. Miletin, "Assessment of neonatal heart rate immediately after birth using digital stethoscope, handheld ultrasound and electrocardiography: An observational cohort study," *Arch. Disease Childhood Fetal Neonatal Ed.*, vol. 104, no. 2, p. F227, Mar. 2019.
- [9] J. M. Hill, A. Maloney, K. Stephens, and L. Eisenfeld, "Stethoscope for monitoring neonatal abdominal sounds," in *Proc. IAJC-IJME Int. Conf.*, 2018, Paper 092.
- [10] V. D. Gaertner, A. C. Kevat, P. G. Davis, and C. O. F. Kamlin, "Evaluation of a digital stethoscope in transitioning term infants after birth," *Arch. Disease Childhood Fetal Neonatal Ed.*, vol. 102, no. 4, pp. F370–F371, Jul. 2017.
- [11] A. C. Kevat, J. Dawson, P. G. Davis, and C. O. F. Kamlin, "Evaluation of a digital stethoscope and smart device technology for assessment of heart rate in the newborn infant," *Arch. Disease Childhood Fetal Neonatal Ed.*, vol. 100, no. 6, pp. F562–F563, Nov. 2015.
- [12] L. Zhou, F. Marzbanrad, A. Ramanathan, D. Fattahi, P. Pharande, and A. Malhotra, "Acoustic analysis of neonatal breath sounds using digital stethoscope technology," *Pediatric Pulmonol.*, vol. 55, no. 3, p. 624, 2020.
- [13] A. Ramanathan, F. Marzbanrad, K. Tan, F.-T. Zohra, M. Acchiardi, R. Roseby, A. Kevat, and A. Malhotra, "Assessment of breath sounds at birth using digital stethoscope technology," *Eur. J. Pediatrics*, vol. 179, pp. 1–9, May 2020.
- [14] M. El-Segaier, O. Lilja, S. Lukkarinen, L. Sörnmo, R. Sepponen, and E. Pesonen, "Computer-based detection and analysis of heart sound and murmur," *Ann. Biomed. Eng.*, vol. 33, no. 7, pp. 937–942, Jul. 2005.
- [15] M. El-Segaier, E. Pesonen, S. Lukkarinen, K. Peters, L. Sörnmo, and R. Sepponen, "Detection of cardiac pathology: Time intervals and spectral analysis," *Acta Paediatrica*, vol. 96, no. 7, pp. 1036–1042, May 2007.
- [16] J. Oliveira, F. Renna, T. Mantadelis, and M. Coimbra, "Adaptive sojourn time HSMM for heart sound segmentation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 642–649, Mar. 2019.
- [17] J. P. de Vos and M. M. Blanckenberg, "Automated pediatric cardiac auscultation," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 2, pp. 244–252, Feb. 2007.
- [18] D. Emmanouilidou, E. D. Mccollum, D. E. Park, and M. Elhilali, "Computerized lung sound screening for pediatric auscultation in noisy field environments," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 7, pp. 1564–1574, Jul. 2018.
- [19] A. M. Amiri and G. Armano, "An intelligent diagnostic system for congenital heart defects," *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 7, pp. 93–98, 2013.
- [20] A. M. Amiri, M. Abtahi, N. Constant, and K. Mankodiya, "Mobile phonocardiogram diagnosis in newborns using support vector machine," *Healthcare*, vol. 5, no. 1, p. 16, 2017.
- [21] S. R. Thiagaraja, R. Dantu, P. L. Shrestha, A. Chitnis, M. A. Thompson, P. T. Anumandla, T. Sarma, and S. Dantu, "A novel heart-mobile interface for detection and classification of heart sounds," *Biomed. Signal Process. Control*, vol. 45, pp. 313–324, 2018.
- [22] A. Djebbari and F. B. Reguig, "Short-time Fourier transform analysis of the phonocardiogram signal," in *Proc. 7th IEEE Int. Conf. Electron., Circuits Syst. (ICECS)*, vol. 2, Dec. 2000, pp. 844–847.
- [23] D. Kumar, P. Carvalho, M. Antunes, R. P. Paiva, and J. Henriques, "Noise detection during heart sound recording using periodicity signatures," *Physiol. Meas.*, vol. 32, no. 5, pp. 599–618, May 2011.
- [24] H. Ghassemian and A. R. Kenari, "Early detection of pediatric heart disease by automated spectral analysis of phonocardiogram in children," *J. Inf. Syst. Telecommun. (JIST)*, vol. 3, no. 2, pp. 66–74, 2015.
- [25] M. Ghaffari, M. Ashourian, E. A. Ince, and H. Demirel, "Phonocardiography signal processing for automatic diagnosis of ventricular septal defect in newborns and children," in *Proc. 9th Int. Conf. Comput. Intell. Commun. Netw. (CICN)*, Sep. 2017, pp. 62–66.
- [26] K. Shirota, "Acoustic analysis of neonatal breath sounds: Changes during the breath adaptation period," *J. Jpn. Bronchoesophagol. Soc.*, vol. 49, no. 5, pp. 442–450, 1998.
- [27] J. F. Kanga and S. S. Kraman, "Comparison of the lung sound frequency spectra of infants and adults," *Pediatric Pulmonol.*, vol. 2, no. 5, pp. 292–295, 1986, doi: 10.1002/ppul.1950020508.
- [28] H. Pasterkamp, R. E. Powell, and I. Sanchez, "Lung sound spectra at standardized air flow in normal infants, children, and adults," *Amer. J. Respiratory Crit. Care Med.*, vol. 154, no. 2, pp. 424–430, 1996.
- [29] T. Imai, "Frequency spectral analysis of breath sounds in newborn infants," *Jpn. J. Pediatric Pulmonol.*, vol. 4, no. 1, pp. 18–27, 1993.
- [30] N. Gavriely, M. Nissan, A. H. E. Rubin, and D. W. Cugell, "Spectral characteristics of chest wall breath sounds in normal subjects," *Thorax*, vol. 50, no. 12, pp. 1292–1300, Dec. 1995.
- [31] L. Vannuccini, J. E. Earis, P. Helistö, B. M. G. Cheetham, M. Rossi, A. R. A. Sovijärvi, and J. Vanderschoot, "Standardization of computerized respiratory sound analysis: Capturing and preprocessing of respiratory sounds," *Eur. Respiratory Rev.*, vol. 10, no. 77, pp. 616–620, 2000.
- [32] G. Charbonneau, E. Ademovic, B. M. G. Cheetham, L. P. Malmberg, J. Vanderschoot, and A. R. A. Sovijärvi, "Standardization of computerized respiratory sound analysis: Basic techniques for respiratory sound analysis," *Eur. Respiratory Rev.*, vol. 10, no. 77, pp. 625–635, 2000.
- [33] M. Bahoura, "Analyse des signaux acoustiques respiratoires: Contribution à la détection automatique des sibilants par paquets d'ondelettes," Ph.D. dissertation, Rouen Univ., Mont-Saint-Aignan, France, 1999.
- [34] S. Reichert, R. Gass, C. Brandt, and E. Andres, "Analysis of respiratory sounds: State of the art," *Clin. Med. Circulatory, Respiratory Pulmonary Med.*, vol. 2, pp. 45–58, May 2008.
- [35] P. He, T. She, W. Li, and W. Yuan, "Single channel blind source separation on the instantaneous mixed signal of multiple dynamic sources," *Mech. Syst. Signal Process.*, vol. 113, pp. 22–35, Dec. 2018.
- [36] S. L. Vo Dinh Minh, "PCA-based human auditory filter bank for speech recognition," in *Proc. Int. Conf. Signal Process. Commun. (SPCOM)*, Dec. 2004, pp. 393–397.
- [37] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Trans. Audio, Speech, Language Process.*, vol. 15, no. 3, pp. 1066–1074, Mar. 2007.

- [38] M. Samieinasab and R. Sameni, "Fetal phonocardiogram extraction using single channel blind source separation," in *Proc. 23rd Iranian Conf. Electr. Eng.*, May 2015, pp. 78–83.
- [39] C. Lin and E. Hasting, "Blind source separation of heart and lung sounds based on nonnegative matrix factorization," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Nov. 2013, pp. 731–736.
- [40] A. Mondal, P. Bhattacharya, and G. Saha, "Reduction of heart sound interference from lung sound signals using empirical mode decomposition technique," *J. Med. Eng. Technol.*, vol. 35, nos. 6–7, pp. 344–353, 2011, doi: [10.3109/03091902.2011.595529](https://doi.org/10.3109/03091902.2011.595529).
- [41] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, 2009, doi: [10.1142/S1793536909000047](https://doi.org/10.1142/S1793536909000047).
- [42] M. A. Colominas, G. Schlotthauer, and M. E. Torres, "Improved complete ensemble EMD: A suitable tool for biomedical signal processing," *Biomed. Signal Process. Control*, vol. 14, pp. 19–29, Nov. 2014.
- [43] F. Ghaderi, H. R. Mohseni, and S. Sanei, "Localizing heart sounds in respiratory signals using singular spectrum analysis," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 12, pp. 3360–3367, Dec. 2011.
- [44] A. Zeiler, R. Faltermeier, I. R. Keck, A. M. Tomé, C. G. Puntonet, and E. W. Lang, "Empirical Mode Decomposition—an introduction," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2010, pp. 1–8.
- [45] B. Mijović, M. De Vos, I. Gligorićević, J. Taelman, and S. Van Huffel, "Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 9, pp. 2188–2196, Sep. 2010.
- [46] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proc. F, Radar Signal Process.*, vol. 140, no. 6, pp. 362–370, 1993.
- [47] L. Tong, V. Soon, Y. Huang, and R. Liu, "AMUSE: A new blind identification algorithm," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 3, May 1990, pp. 1784–1787.
- [48] R. Sameni, C. Jutten, and M. B. Shamsollahi, "Multichannel electrocardiogram decomposition using periodic component analysis," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1935–1940, Aug. 2008.
- [49] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.
- [50] J. Miettinen, K. Nordhausen, and S. Taskinen, "Blind source separation based on joint diagonalization in R: The packages JADE and BSSasymp," *J. Stat. Softw.*, vol. 76, no. 2, pp. 1–31, 2017.
- [51] J. E. Fowler, "The redundant discrete wavelet transform and additive noise," *IEEE Signal Process. Lett.*, vol. 12, no. 9, pp. 629–632, Sep. 2005.
- [52] J. E. Fowler, "The redundant discrete wavelet transform and additive noise," *IEEE Signal Process. Lett.*, vol. 12, no. 9, pp. 629–632, Sep. 2005.
- [53] S. Ulukaya, G. Serbes, and Y. P. Kahya, "Wheeze type classification using non-dyadic wavelet transform based optimal energy ratio technique," *Comput. Biol. Med.*, vol. 104, pp. 175–182, Jan. 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482518303421>
- [54] S. Ulukaya, G. Serbes, and Y. P. Kahya, "Overcomplete discrete wavelet transform based respiratory sound discrimination with feature and decision level fusion," *Biomed. Signal Process. Control*, vol. 38, pp. 322–336, Sep. 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809417301325>
- [55] B. Ergen, Y. Tatar, and H. O. Gulcur, "Time–frequency analysis of phonocardiogram signals using wavelet transform: A comparative study," *Comput. Methods Biomechanics Biomed. Eng.*, vol. 15, no. 4, pp. 371–381, Apr. 2012.
- [56] Z. M. Zin, S. H. Salleh, S. Daliman, and M. D. Sulaiman, "Analysis of heart sounds based on continuous wavelet transform," in *Proc. Student Conf. Res. Develop. (SCORED)*, 2003, pp. 19–22.
- [57] U. K. Tapmedia Ltd, London. *VoiceRecorder*. Accessed: Dec. 31, 2019. [Online]. Available: <https://www.tapmedia.co.U.K>
- [58] E. Grooby, J. He, J. Kiewsky, D. Fattahi, L. Zhou, A. King, A. Ramanathan, A. Malhotra, G. A. Dumont, and F. Marzbanrad, "Neonatal heart and lung sound quality assessment for robust heart and breathing rate estimation for telehealth applications," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 12, pp. 4255–4266, Dec. 2021.
- [59] I. Daubechies, *Discrete Wavelet Transforms: Frames* (Society for Industrial and Applied Mathematics). Philadelphia, PA, USA: SIAM, 1992, pp. 53–105, ch. 3.
- [60] I. Daubechies, *Time-Frequency Density and Orthonormal Bases* (Society for Industrial and Applied Mathematics). Philadelphia, PA, USA: SIAM, 1992, pp. 107–128, ch. 4.
- [61] G. Shah, P. Koch, and C. B. Papadias, "On the blind recovery of cardiac and respiratory sounds," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 1, pp. 151–157, Jan. 2015.
- [62] J. Le Roux, F. Weninger, and J. Hershey, "Sparse NMF—Half-baked or well done?" Mitsubishi Electr. Res. Laboratories, Cambridge MA, USA, Tech. Rep. TR2015-023, Mar. 2015. [Online]. Available: <http://www.jonathanleroux.org/software>. <https://www.merl.com/publications/TR2015-023>
- [63] M. A. Colominas. (Feb. 2015). *A MATLAB Package for CEEMDAN (Complete Ensemble Empirical Mode Decomposition With Adaptive Noise)*. [Online]. Available: <https://github.com/macolominas/CEEMDAN>
- [64] G. Rilling and P. Flandrin. (Mar. 2007). *MATLAB/C Codes for EMD and EEMD With Examples*. [Online]. Available: <https://github.com/macolominas/CEEMDAN>
- [65] B. Gao. (Dec. 2014). *Cochleagram and IS-NMF2D for Blind Source Separation*. (Jan. 9, 2022). [Online]. Available: <https://www.mathworks.com/matlabcentral/fileexchange/48622-cochleagram-%and-is-nmf2d-for-blind-source-separation>
- [66] R. Sameni. (2018). *The Open-Source Electrophysiological Toolbox (OSET), Version 3.14*. [Online]. Available: <http://www.uset.ir>
- [67] E. Grooby, C. Sitaula, D. Fattahi, R. Sameni, K. Tan, L. Zhou, A. King, A. Ramanathan, A. Malhotra, G. A. Dumont, and F. Marzbanrad, "Real-time multi-level neonatal heart and lung sound quality assessment for telehealth applications," *IEEE Access*, vol. 10, pp. 10934–10948, 2022, doi: [10.1109/ACCESS.2022.3144355](https://doi.org/10.1109/ACCESS.2022.3144355).
- [68] D. B. Springer, L. Tarassenko, and G. D. Clifford, "Logistic regression-HSMM-based heart sound segmentation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 822–832, Sep. 2016.
- [69] A. Kazemnejad, P. Gordany, and R. Sameni, "An open—Access simultaneous electrocardiogram and phonocardiogram database," *bioRxiv*, May 2021.
- [70] A. Kazemnejad, P. Gordany, and R. Sameni. (2021). *EPHNOGRAM: A Simultaneous Electrocardiogram and Phonocardiogram Database*. [Online]. Available: <https://physionet.org/content/ephnoqram/1.0.0/>



**DAVOOD FATTAHI** is currently pursuing the Ph.D. degree in biomedical engineering (bio-electric) with Shiraz University, Shiraz, Iran. He spent a Research Internship (March 2019–September 2019) with the Biomedical Signal Processing Laboratory, Department of Electrical and Computer Systems Engineering, Monash University, Australia. His research interests include biomedical signal processing (ECG, EEG, PCG, lung sound), parameter estimation, time-frequency analysis, and blind source separation.



**REZA SAMENI** (Senior Member, IEEE) received the bachelor's degree in electronics engineering from Shiraz University, Iran, in 2000, the master's degree in biomedical engineering from the Sharif University of Technology, Iran, in 2003, and the double Ph.D. degree in signal processing and biomedical engineering from the Institut National Polytechnique de Grenoble (INPG), France, and the Sharif University of Technology, in 2008. He was a tenured Associate Professor

with the School of Electrical and Computer Engineering, Shiraz University, from 2008 to 2018, an Invited Senior Researcher with the GIPSA-Laboratory, Grenoble, France, from 2018 to 2020. He has been an Associate Professor in biomedical engineering at Emory University, GA, USA, since 2020. His research interests include statistical signal processing with special interests in mathematical modeling and analysis of biomedical systems and signals.



**ETHAN GROOBY** (Member, IEEE) received the bachelor’s degree in biomedicine and the Master of Engineering degree in biomedical from The University of Melbourne, Melbourne, VIC, Australia, in 2017 and 2019, respectively. He is currently pursuing the joint Ph.D. degree in electrical and computer systems engineering with Monash University, Melbourne, and The University of British Columbia, Vancouver, BC, Canada. From 2017 to 2019, he was a Research Student with the Peter MacCallum Cancer Centre, Walter and Eliza Hall Institute of Medical Research, and the Defence Science and Technology Group. In 2019 and 2020, he was a Research Engineer at Cochlear. His research interests include biomedical signal processing and medical device development.



**ASHWIN RAMANATHAN** received the bachelor’s degree (Hons.) in medical science and the M.B.B.S. degree (Hons.) from the Department of Paediatrics, Monash University, Melbourne, Australia. He is currently a Resident Medical Officer at Perth Children’s Hospital.



**KENNETH TAN** is currently a Consultant Neonatologist at Monash Children’s Hospital and an Adjunct Associate Professor at Monash University. His research interests include clinical registry, clinical trials, big data analysis, and clinical practice improvement.



**ATUL MALHOTRA** received the Ph.D. and M.D. degrees. He is also a Senior Neonatologist at Monash Children’s Hospital, and an Associate Professor (a Research)/a NHMRC Fellow at Monash University, Melbourne, Australia. He has a large research program, with interests in neonatal lung and brain injury, with more than \$7 million in research funding. He has published more than 100 peer-reviewed articles, and four book chapters to date. Together with Dr. Fae Marzbanrad, their

team researches digital health technologies to improve neonatal cardiorespiratory monitoring.



**LINDSAY ZHOU** received the M.B.B.S. degree, in 2012. He is currently a Neonatologist at Monash Children’s Hospital, Melbourne, Australia, completing his neonatal fellowship training, in 2020. He is undertaking a Ph.D. investigating umbilical cord blood-derived cell therapies for preterm brain injury at Monash University.



**ARRABELLA KING** received the bachelor’s degree (Hons.) in medical science, in 2019, and the bachelor’s degree in medical science and the Doctor degree in medicine from Monash University, Melbourne, Australia, in 2020. She is currently working as a Junior Doctor at St. Vincent’s Hospital, Melbourne, and holds an affiliate position with the Department of Paediatrics, Monash University.



**FAEZEH MARZBANRAD** (Senior Member, IEEE) received the Ph.D. degree from The University of Melbourne, Australia, in 2016. She is currently a Lecturer and the Head of the Biomedical Signal Processing Laboratory, Department of Electrical and Computer Systems Engineering, Monash University, Australia. Her research interests include biomedical signal processing, machine learning, affordable medical technologies, and mobile-health.

...