

Received 7 January 2024, accepted 18 February 2024, date of publication 29 February 2024, date of current version 8 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3371510

METHODS

Separation of the Aortic and Pulmonary Components of the Second Heart Sound via Alternating Optimization

FRANCESCO RENNA¹, (Senior Member, IEEE), ALEX GAUDIO², (Member, IEEE), SANDRA MATTOS³, MARK D. PLUMBLEY⁴, (Fellow, IEEE), AND MIGUEL TAVARES COIMBRA¹, (Senior Member, IEEE)

¹INESC TEC, Faculdade de Ciências da Universidade do Porto, 4169-007 Porto, Portugal

²Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA

³Maternal-Fetal Cardiac Unit, Royal Portuguese Hospital, Recife 52010-075, Brazil

⁴Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, GU2 7UX Guildford, U.K.

Corresponding author: Alex Gaudio (agaudio2@jh.edu)

This work was supported in part by the Component 5, Capitalization and Business Innovation, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the Scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the Next Generation EU (2021–2026) within project Health from Portugal (HfPT), with reference 41; and in part by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant number EP/T019751/1.

ABSTRACT An algorithm for blind source separation (BSS) of the second heart sound (S2) into aortic and pulmonary components is proposed. It recovers aortic (A2) and pulmonary (P2) waveforms, as well as their relative delays, by solving an alternating optimization problem on the set of S2 sounds, without the use of auxiliary ECG or respiration phase measurement data. This unsupervised and data-driven approach assumes that the A2 and P2 components maintain the same waveform across heartbeats and that the relative delay between onset of the components varies according to respiration phase. The proposed approach is applied to synthetic heart sounds and to real-world heart sounds from 43 patients. It improves over two state-of-the-art BSS approaches by 10% normalized root mean-squared error in the reconstruction of aortic and pulmonary components using synthetic heart sounds, demonstrates robustness to noise, and recovery of splitting delays. The detection of pulmonary hypertension (PH) in a Brazilian population is demonstrated by training a classifier on three scalar features from the recovered A2 and P2 waveforms, and this yields an auROC of 0.76.

INDEX TERMS Alternating optimization, auscultation, blind source separation, second heart sound.

I. INTRODUCTION

A. MOTIVATION

Cardiovascular diseases are the leading cause of death worldwide [1]. Among the different cardiovascular diseases, those associated with pulmonary hypertension (PH) represent a significant burden for healthcare systems because of difficulties in the early recognition of their symptoms as well as the complexity and cost of the corresponding treatments [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Kin Fong Lei¹.

Direct measurement of pulmonary artery pressure (PAP) enables the diagnosis of PH and heart failure, but a non-invasive gold standard for direct measurement does not currently exist. Direct measurement via Swan-Ganz right heart catheterization or a CardioMEMS sensor implant [3] is invasive, expensive, and only suitable for patients with a high probability of PH. Recent research efforts for non-invasive estimates of pulmonary pressures use Doppler echocardiography. While echocardiography is relatively low cost and has minimal risk, it presents significant limitations for PAP estimation. Echocardiography requires specialized equipment and a trained clinician. PAP estimates cannot

be obtained via echocardiography in approximately 50% of patients with normal PAP, nor in 10–20% of patients with elevated PAP, nor in patients with absence of tricuspid regurgitation [4]. Furthermore, Doppler-based PAP estimates present an average error of 30% when compared with the “gold standard” offered by right heart catheterization [5]. Moreover, patients receive Doppler PAP estimates when they have been referred to echocardiography and these patients typically already present significant symptoms of PH. Cardiac auscultation with a digital stethoscope offers a more timely way to perform PH screening, such as in point-of-care settings where PH exists but is asymptomatic. Digital cardiac auscultation is lower cost test with minimal training requirements and wide applicability.

The second heart sound (S2) in particular can be analyzed to estimate the PAP. The S2 sound is formed by the vibrations that occur as a result of the aortic and pulmonary valve closures, and the subsequent vibrations that occur in the cusps and in the walls and blood columns of the vessels and ventricles. The S2 sound can be modeled as a summation of A2 and P2 components. The A2 component represents sound emitted by the aortic valve closure, and the P2 component represents the pulmonary valve closure. In medical literature, the audible delay between the onset of the A2 and P2 components is known as splitting [6]. The delay increases during inspiration and decreases during exhalation due to respiration’s effect on thoracic pressure [7], and analysis of splitting is clinically relevant for diagnosis of diseases and heart conditions including hypertension. The manual detection of PH via traditional cardiac auscultation is highly dependent on the expertise and subjective evaluation of the operator, where the tasks are to first recognize subtle differences in amplitude and clicks of the A2 and P2 components, and then decide if the spacing and amplitude is out of normal range. Fig. 1 visualizes an example human S2 sound.

Evidence from the literature suggests that accurate estimation of the PAP can benefit by computer methods to separate the A2 and P2 components of the S2 sound [8]. From a signal processing perspective, the separation of A2 and P2 components is challenging due to the overlap of signals in the time and frequency domain, and due to the difficulty in attributing the observed sound to the heart valve that caused it. Moreover, standard methods for blind source separation, such as independent component analysis (ICA) [9] and morphological component analysis (MCA) [10], impose an assumption of independence between components that does not model the conditional dependence of the A2 and P2 components on the splitting delay caused by respiration.

In this work, we propose a novel algorithm for the separation of the A2 and P2 components of S2 heart sounds. The main outputs of the algorithm, visualized in Fig. 1, are the A2 and P2 waveforms and the splitting delay. The main contributions of the paper are:

- 1) The proposal of a novel, non-parametric and unsupervised S2 separation algorithm that uses alternating

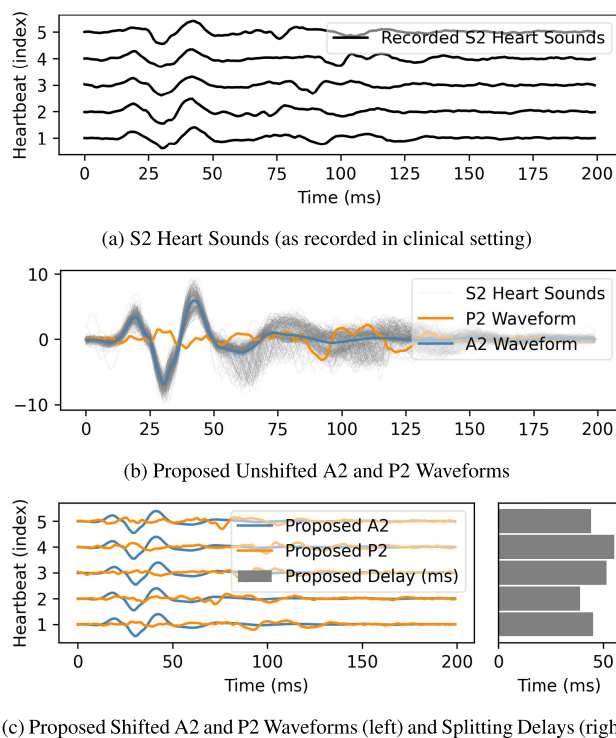


FIGURE 1. The novel unsupervised algorithm analyzes S2 sounds (1a) to recover A2 and P2 waveforms (1c, left) and associated time delay (1c, right). The proposed alternating optimization method obtains a single P2 and A2 waveform for all heartbeats (1b), and the shifted P2 in Fig. 1c. The shifted P2 is obtained as the difference of the A2 waveform from the recorded S2 sound for each heartbeat.

optimization to separate a collection of S2 sounds into a proposed unshifted A2 waveform and unshifted P2 waveform (see Fig. 1b), and a set of splitting delays and corresponding shifted waveforms (see Fig. 1c);

- 2) The analysis of the proposed S2 separation algorithm and its comparison with the state-of-the-art separation approaches on a set of synthetically generated heart sounds;
- 3) The application of the proposed separation algorithm to real-world cardiac auscultation data to detect PH and extract relevant features.

The remainder of this paper is organized as follows. Related works are considered in Section I-B. Section II contains the description of the proposed approach for S2 source separation. Section III empirically validates the separation algorithm on synthetic signals. Section IV applies the proposed method to real-world heart sounds for pulmonary hypertension detection. Finally, the obtained results are discussed in Section V and conclusions are drawn in Section VI.

B. RELATED WORK

We identify four different classes of approaches that have appeared in the literature in recent years to separate S2 sounds into their components: (1) standard BSS algorithms that do not account for specific characteristics of the S2 sound;

(2) *ad hoc*, predetermined waveforms; (3) signal decompositions; and (4) methods based on splitting statistics/dynamics.

A first class of approaches uses standard blind source separation (BSS) algorithms to detect A2 and P2 components of S2 sounds [11], [12]. ICA [9] assumes that the components constituting the observed signal are statistically independent and synchronized in the time domain. These two assumptions do not fit well with the underlying generation of the S2 sound. For instance, the onset of P2 can be inferred through its conditional dependence on the onset of A2 and the respiratory phase. A general mathematical model of shifted ICA [13] can model splitting statistics. In this work we leverage some tools from shifted ICA to simultaneously capture the A2 and P2 waveforms and their delays. Explicit separation of the A2 and P2 components with BSS is an open problem.

A second class of approaches models the A2 and P2 components via an *ad hoc* pre-determined waveform generating function, whose parameters are fitted to the observed S2 data [4], [14], [15], [16], [17], [18], [19]. For instance, transient nonlinear chirp models were developed as a result of analysis of isolated A2 and P2 components extracted from heart sounds recorded in pigs [4], [14], [20]. Information regarding the concentration of the components around their instantaneous frequencies was obtained from the visual inspection of the signal's Wigner-Ville distribution. A fully automatic approach to separating the S2 sound into its components considered Gaussian chirplets [15] fitted to the observed S2 signal using a statistically optimal null filter [16]. Similarly, [17] uses the continuous wavelet transform to decompose the observed S2 sounds using a nonlinear transient chirp as a mother wavelet. The A2 and P2 components have also been modeled with Gabor wavelets [19]. More recently, the A2 and P2 components have been modeled via Gaussian windowed sinusoids, whose parameters (amplitude, location, and width of the Gaussian windows as well as sinusoids frequency and initial phase) are obtained by solving a least-squares optimization problem [18]. By design, the methods in this class of approaches are limited by the need to define a hand-crafted waveform or waveform generating function. Such *ad hoc* approaches are not flexible enough to model the natural variation of S2 sounds.

A third class of approaches uses signal decomposition as a way of retrieving A2 and P2 components [21], [22], [23], [24]. A Fourier transform of manually cropped regions of the S2 was used to analyze A2 and P2 components [25]. The continuous and discrete wavelet transforms have been used to visually represent A2 and P2 components and manually find their relative delay by visual inspection [21], [22]. The Hilbert transform has also been utilized, along with a wavelet transform, for the analysis of the A2 and P2 splitting delay [24]. A slightly more automated method for detecting peaks of the A2 and P2 components and their relative delay uses the Hilbert vibration decomposition (HVD) followed by an analysis of the smoothed pseudo Wigner-Ville distribution [23]. However, this work still uses

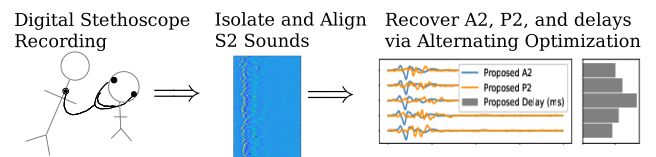


FIGURE 2. Application to real-world data: Recovery of A2 and P2 components and splitting statistics via an unsupervised alternating optimization algorithm. Our main contribution is the alternating optimization step.

visual inspection to choose a number of components retained from the HVD.

A fourth class of approaches has recently emerged to provide fully automatic separation of S2 into A2 and P2 components, via analysis of splitting delays due to respiration. In particular, the work by Tang et al. [26] exploits an assumption that, across many heartbeats, the A2 component occupies the same position in the S2, whereas P2 component experiences a relative shift according to the respiration phase. Therefore, the A2 components are estimated by averaging the collected S2 sounds and the P2 components are obtained by subtracting the derived A2 component from the S2 sound. Secondly, the assumption that A2 is obtained by averaging is a restrictive assumption that makes the algorithm sensitive to alignment of S2 sounds. Other work attempts to recover the A2 and P2 components using a joint multivariate Gaussian mixture model (GMM) from S2 sounds [27]. However, this approach is a supervised machine learning approach that requires the availability of ground truth annotated A2 and P2 components. An annotated dataset of A2 and P2 sounds is difficult or impossible to collect on human subjects or in standard clinical practice because it is not clear how exactly to separate the superimposed A2 and P2 components, and the model was not tested on real-world data.

The proposed blind source separation algorithm overcomes shortcomings of the surveyed literature by extending previous preliminary results presented in [28] to separate the A2 and P2 components of the S2. While *ad hoc* pre-determined waveform restrictions limit the utility of some of the approaches previously presented in the literature, the proposed method does not pre-define the waveform. Some approaches study splitting statistics to obtain A2 and P2, but they require visual inspection and manual human intervention. In contrast, the proposed method is fully automated. Some of the fully automated approaches suffer from the lack of available data; the proposed method is unsupervised and requires only S2 heart sound data.

II. PROPOSED METHOD

The proposed algorithm recovers the aortic (A2) and pulmonary (P2) components as well as their relative delays from a set of S2 heart sounds. The overall approach and its underlying modeling assumptions are discussed in detail. Fig. 2 visually summarizes the approach applied to

Algorithm 1 Proposed S2 Sound Source Separation Algorithm

Require: $\{s_m(tT_s) \mid m \in \{1..M\}, t \in \{0..N-1\}\}$ \triangleright A set of M S2 sound samples, each of length N , previously extracted from a PCG recording \triangleright

- 1: Initialize A2 component estimates (See Eq. 3) \triangleright
 $\hat{a}^{(0)}(tT_s) = \frac{1}{M} \sum_m s_m(tT_s)$
 $\hat{\tau}_{A,m}^{(0)} = 0 \forall m$ \triangleright
- 2: Initialize P2 component estimates (See Eq. 4, 5, and 6) \triangleright
 $\hat{p}^{(0)}(tT_s) = \frac{1}{M} \sum_m s_m(tT_s) - \hat{a}^{(0)}(tT_s - \hat{\tau}_{P,m}^{(0)})$
 $\hat{\tau}_{P,m}^{(0)} = \arg \max_{tT_s} \rho_1^{(0)} \otimes \rho_m^{(0)} \quad \forall m = 1, \dots, M$ \triangleright
- 3: Implement Iterative Alternating Optimization
- 4: **for** $i \in \{1..I\}$ **do** \triangleright
- 5: Estimate Delays via Projected Gradient Descent (see Eq. 7 and 10) \triangleright
 $[\hat{\tau}_{A,m}^{(i)}, \hat{\tau}_{P,m}^{(i)}] = \arg \min_{\tau_{A,m}, \tau_{P,m}} \sum_f \left| S_m(fF_s) - \hat{A}^{(i)}(fF_s) e^{-j2\pi f F_s \tau_{A,m}} \right|$ \triangleright
- 6: Estimate A2 and P2 Waveforms (via Moore-Penrose pseudoinverse, see Eq. 11–15) \triangleright
 $[\hat{a}^{(i+1)}(\cdot), \hat{p}^{(i+1)}(\cdot)] = \arg \min_{a(\cdot), p(\cdot)} \sum_m \sum_t \left| s_m(tT_s) - a(tT_s - \hat{\tau}_{A,m}^{(i)}) - p(tT_s - \hat{\tau}_{P,m}^{(i)}) \right|^2$ \triangleright
- 7: **end for** \triangleright
- 8: Obtain A2 and P2 waveforms (see Eq. 16 and 17) \triangleright
 $a(tT_s - \tau_{A,m})^* = \hat{a}^{(I)}(tT_s - \hat{\tau}_{A,m})$
 $p(tT_s - \tau_{P,m})^* = s_m(tT_s) - \hat{a}^{(I)}(tT_s - \hat{\tau}_{A,m}^{(I)})$ \triangleright
- 9: Return waveforms and delays \triangleright
return $a(\cdot)^*, p(\cdot)^*, \{\hat{\tau}_{A,m}^{(I)}\}, \{\hat{\tau}_{P,m}^{(I)}\}$

real-world data, and Algo. 1 presents the proposed method as an algorithm.

A. SIGNAL MODEL AND SETUP

We are presented with the observation of a series of M S2 sounds obtained from a single heart sound recording of a given patient. It is assumed that suitable pre-processing has been applied to first extract and align the S2 sounds in the heart sound audio recording. An example preprocessing approach is described in Section IV-D. The observed S2 signals are assumed to be sampled with period T_s . Moreover, N samples are collected for each S2 sound, where N depends on the number of heartbeats in the recording. For example, Fig. 1 depicts $M = 5$ recorded S2 sounds, each of $N = 200$ samples, and sampling period of $T_s = 1$ millisecond.

Assumptions and Motivation: It is assumed that each observed S2 sound can be described as a sum of recovered A2 and P2 components, or informally $s(\cdot) = a(\cdot) + p(\cdot)$. As a blind method, it is assumed that no prior information is disclosed about the shape of the waveforms $a(\cdot)$ and $p(\cdot)$. It is also assumed that the stethoscope and subject do not move relative to each other during recording and that the volume of the S2 signal is approximately unchanged

during recording. The model used to recover A2 and P2 components is motivated by the observation that respiratory activity modulates the morphology of S2 sounds through changes in the pleural pressure and the pulmonary blood flow. Namely, at late inspiration and early expiration, the relative time distance (i.e., the split) between the A2 and P2 components is increased due to the elevated pressure in the right ventricle, whereas the morphology of different S2 sounds is kept approximately constant during apnea¹ [7]. For this reason, we assume that the A2 and P2 components in different S2 sounds of the same recording, respectively, have the same shape but different delays. The observed S2 sounds are modeled as follows:

$$s_m(tT_s) = a(tT_s - \tau_{A,m}) + p(tT_s - \tau_{P,m}), \quad (1)$$

where $m = 1, \dots, M$ indexes the heart sounds of the patient, $t = 0, \dots, N-1$ indexes samples of the signal, and the $\tau_{\cdot,m}$ terms identify delays (offsets) from the start of the S2 sound.

Objective: The objective of the considered source separation method is to recover the waveforms $a(\cdot)$ and $p(\cdot)$ and the delays $\tau_{A,m}$ and $\tau_{P,m}$ from the observation of the S2 signal $s_m(tT_s)$, for all $m = 1, \dots, M$ observed S2 sounds.

The method proposed for separating the A2 and the P2 components of the recorded S2 sounds formulates a least-squares optimization problem whose solution is approximated via an alternating iterative approach. The considered optimization problem is expressed as follows:

$$\begin{aligned} & [a(\cdot), p(\cdot), \{\tau_{A,m}\}, \{\tau_{P,m}\}]^* \\ & = \arg \min_{a(\cdot), p(\cdot), \{\tau_{A,m}\}, \{\tau_{P,m}\}} \sum_{m=0}^{M-1} \sum_{t=0}^{N-1} |s_m(tT_s) \\ & \quad - a(tT_s - \tau_{A,m}) - p(tT_s - \tau_{P,m})|^2, \end{aligned} \quad (2)$$

which aims to minimize the mean-squared error (MSE) between the S2 observations and the superposition of the separated A2 and P2 components. This objective assumes that the A2 and P2 sounds in different heartbeats have the same respective shapes but different delays.

After an initialization step, the optimal solution (2) will be approximated by an alternating iterative algorithm which alternates between finding the sets of delays $\{\hat{\tau}_{A,m}\}, \{\hat{\tau}_{P,m}\}$ that minimize the MSE with a given pair of waveforms $\hat{a}(\cdot), \hat{p}(\cdot)$, and alternately finding the waveforms $\hat{a}(\cdot), \hat{p}(\cdot)$ that minimize the MSE for given sets of delays $\{\hat{\tau}_{A,m}\}, \{\hat{\tau}_{P,m}\}$. As it will be shown in the remainder of this section, the proposed alternating optimization approach simplifies the problem in (2) and outputs an inferred shape of the waveforms associated with the A2 and P2 components as well as their corresponding delays in each heartbeat.

¹Apnea is a medical term that describes the temporal cessation of breathing, such as between inhale and exhale, when there is no movement of the muscles and the volume of the lungs is not changing.

B. INITIALIZATION

The initial estimates of waveforms and delays associated to A2 and P2 components are obtained by applying the method described in [26], which computes an estimate of the A2 component by simply averaging the observed S2 sounds. This method assumes the A2 components appear with the same delay in each observed S2 sound, while the contributions of the corresponding P2 components have different delays in each S2 sound, and are zeroed out by averaging. Therefore, we can write the initial estimate $\hat{a}^{(0)}(tT_s)$ as:

$$\hat{a}^{(0)}(tT_s) = \frac{1}{M} \sum_m s_m(tT_s), \tag{3}$$

where we have introduced the use of the superscript $(\cdot)^{(i)}$ to denote the estimate at the i -th iteration of the algorithm. According to the model used in [26], the delays associated to the A2 components are the same in all S2 sounds. Therefore, without loss of generality, we can assume in our initialization that $\hat{\tau}_{A,m}^{(0)} = 0$ for all $m = 1, \dots, M$.

The initial values $\hat{p}^{(0)}(\cdot)$ and $\hat{\tau}_{P,m}^{(0)}$ are obtained by a heuristic approach in three steps. First, an unaligned P2 representation is obtained by subtraction in Eq. (4) for all M delays. Second, delays $\hat{\tau}_{P,m}^{(0)}$ are obtained in Eq. (5) by finding the position in time of the maximum cross correlation between the P2 representations of the first heartbeat ($m = 1$) and each of the M heartbeats. Third, the P2 waveform is obtained in Eq. (6) by averaging an aligned P2 representation:

$$\rho_m^{(0)}(tT_s) = s_m(tT_s) - \hat{a}^{(0)}(tT_s) \tag{4}$$

$$\hat{\tau}_{P,m}^{(0)} = \arg \max_{tT_s} \rho_1^{(0)}(tT_s) \otimes \rho_m^{(0)}(tT_s) \tag{5}$$

$$\hat{p}^{(0)}(tT_s) = \frac{1}{M} \sum_m s_m(tT_s) - \hat{a}^{(0)}(tT_s - \hat{\tau}_{P,m}^{(0)}), \tag{6}$$

where \otimes is the cross correlation operator, and where Eq. (4) and Eq. (5) are evaluated for all $m \in \{1, \dots, M\}$.

C. DELAY ESTIMATION

Given the estimates $\hat{a}^{(i)}(\cdot)$, $\hat{p}^{(i)}(\cdot)$ of the A2 and P2 waveforms obtained from the i -th iteration of the algorithm, the corresponding delay estimates $\{\hat{\tau}_{A,m}^{(i)}\}$, $\{\hat{\tau}_{P,m}^{(i)}\}$ are obtained as the solution of the following problem:

$$\begin{aligned} & \{ \{\hat{\tau}_{A,m}^{(i)}\}, \{\hat{\tau}_{P,m}^{(i)}\} \} \\ &= \arg \min_{\{\tau_{A,m}\}, \{\tau_{P,m}\}} \sum_m \sum_t |s_m(tT_s) \\ & \quad - \hat{a}^{(i)}(tT_s - \tau_{A,m}) - \hat{p}^{(i)}(tT_s - \tau_{P,m})|^2, \end{aligned} \tag{7}$$

where optimization can be performed independently for the different indexes m .

In addition, it is possible to express an optimization problem equivalent to (7) in the frequency domain, by leveraging Parseval’s identity, which says that the integral of the square of a function in the time domain equals the sum of squares of the function’s Fourier coefficients. We denote with

$S_m(fF_s)$, $\hat{A}^{(i)}(fF_s)$, and $\hat{P}^{(i)}(fF_s)$ the discrete Fourier transform (DFT) of time-domain signals $s_m(tT_s)$, $\hat{a}^{(i)}(tT_s)$, and $\hat{p}^{(i)}(tT_s)$, respectively, where $F_s = \frac{1}{N_{\text{DFT}}T_s}$ and $N_{\text{DFT}} \geq N$ is the number of data points used to compute the DFT. Then, the frequency-domain delay optimization problem for a given index m can be written as:

$$[\hat{\tau}_{A,m}^{(i)}, \hat{\tau}_{P,m}^{(i)}] = \arg \min_{\tau_{A,m}, \tau_{P,m}} \sum_f \tag{8}$$

$$\left| S_m(fF_s) - \hat{A}^{(i)}(fF_s) e^{-j2\pi fF_s \tau_{A,m}} \right. \tag{9}$$

$$\left. - \hat{P}^{(i)}(fF_s) e^{-j2\pi fF_s \tau_{P,m}} \right|^2, \tag{10}$$

where the gradient with respect to $[\hat{\tau}_{A,m}^{(i)}, \hat{\tau}_{P,m}^{(i)}]$ can be expressed in closed form. Based on this observation, the delay estimates $\{\hat{\tau}_{A,m}^{(i)}\}$, $\{\hat{\tau}_{P,m}^{(i)}\}$ are obtained by approximating the solution of (10) via projected gradient descent. At each gradient descent iteration, the delays are projected to a physiologically feasible range. The time split between A2 and P2 components usually ranges between 10 ms and 50 ms [26], and we therefore constrain delay values in the interval $[-100, 100]$ ms.

D. WAVEFORM ESTIMATION

Given the current delay estimates $\hat{\tau}_{A,m}^{(i)}$, $\hat{\tau}_{P,m}^{(i)}$, and the observations $s_m(tT_s)$, the signal waveforms associated to the $(i + 1)$ -th iteration, $\hat{a}^{(i+1)}(\cdot)$, $\hat{p}^{(i+1)}(\cdot)$ are obtained again by minimizing the MSE, that is, they are given by the solution the following optimization problem:

$$\begin{aligned} & [\hat{a}^{(i+1)}(\cdot), \hat{p}^{(i+1)}(\cdot)] = \arg \min_{a(\cdot), p(\cdot)} \sum_m \sum_t |s_m(tT_s) \\ & \quad - a(tT_s - \hat{\tau}_{A,m}^{(i)}) - p(tT_s - \hat{\tau}_{P,m}^{(i)})|^2. \end{aligned} \tag{11}$$

Assuming that the admissible delay estimates are always obtained as multiples of the sampling period T_s , rather than interpolated between samples, the problem in (11) can be cast as a standard linear least-squares problem, which is overdetermined when $M > 2$. The problem can be expressed in matrix form as follows:

$$\begin{bmatrix} \mathbf{a}^{(i+1)} \\ \mathbf{p}^{(i+1)} \end{bmatrix} = \arg \min_{\mathbf{a}, \mathbf{p} \in \mathbb{R}^N} \left\| \mathbf{s} - \mathbf{X}^{(i)} \begin{bmatrix} \mathbf{a} \\ \mathbf{p} \end{bmatrix} \right\|_2^2, \tag{12}$$

where the column vector $\mathbf{s} \in \mathbb{R}^{M \cdot N}$ concatenates all N samples of all M observed S2 sound signals of a given recorded signal, the column vector $\mathbf{a} \in \mathbb{R}^N$ contains the samples of the estimated signal $a(\cdot)$, and the column vector $\mathbf{p} \in \mathbb{R}^N$ contains the samples of the estimated signal $p(\cdot)$. The binary “shift” matrix $\mathbf{X}^{(i)} \in \{0, 1\}^{M \cdot N \times 2N}$ is formed by $M \times 2$ blocks of dimension $N \times N$, each containing a “shifted version” of the $N \times N$ identity matrix, where the values of such shifts depend on the delays $\hat{\tau}_{A,m}^{(i)}$ and $\hat{\tau}_{P,m}^{(i)}$.

Namely, we can write

$$\mathbf{X}^{(i)} = \begin{bmatrix} \mathbf{T}_{A,0}^{(i)} & \mathbf{T}_{P,0}^{(i)} \\ \vdots & \vdots \\ \mathbf{T}_{A,m}^{(i)} & \mathbf{T}_{P,m}^{(i)} \\ \vdots & \vdots \\ \mathbf{T}_{A,M}^{(i)} & \mathbf{T}_{P,M}^{(i)} \end{bmatrix}, \quad (13)$$

where $\mathbf{T}_{j,m}^{(i)} \in \{0, 1\}^{N \times N}$ for $j = \{A, P\}$ and the element in the k -th row and ℓ -th column of the matrix $\mathbf{T}_{j,m}^{(i)}$ is given by

$$\mathbf{T}_{j,m}^{(i)}[k, \ell] = \begin{cases} \delta_{k, \ell + \lfloor \hat{\tau}_{j,m}^{(i)} / T_s \rfloor}, & \text{if } \hat{\tau}_{j,m}^{(i)} \geq 0 \\ \delta_{k + \lfloor \hat{\tau}_{j,m}^{(i)} / T_s \rfloor, \ell}, & \text{otherwise,} \end{cases} \quad (14)$$

for $j = \{A, P\}$ and $m = 0, \dots, M - 1$, where $\delta_{k,\ell}$ is the Kronecker delta symbol.

When the matrix $\mathbf{X}^{(i)}$ has full column rank, the corresponding linear least-squares problem (12) has a unique solution which is given by

$$\begin{bmatrix} \mathbf{a}^{(i+1)} \\ \mathbf{p}^{(i+1)} \end{bmatrix} = (\mathbf{X}^{(i)})^\dagger \mathbf{s} = \left((\mathbf{X}^{(i)})^T \mathbf{X}^{(i)} \right)^{-1} (\mathbf{X}^{(i)})^T \mathbf{s}, \quad (15)$$

where $(\cdot)^\dagger$ represents the Moore-Penrose pseudoinverse. After I steps, the obtained unshifted waveforms \mathbf{a}^I and \mathbf{p}^I , such as those shown in Fig. 1b, are subsequently obtained in shifted form in the next sub-section.

E. TERMINATION

The algorithm is terminated after a fixed number I of iterations of alternating delay and waveform estimation processes. After the last iteration, the recovered waveforms \mathbf{a}^I and \mathbf{p}^I are combined with the delays $[\tau_{A,m}]$ and $[\tau_{P,m}]$ to obtain one shifted waveform for each delay:

$$a(tT_s - \tau_{A,m})^* = \hat{a}^{(I)}(tT_s - \hat{\tau}_{A,m}^{(I)}) \quad (16)$$

$$p(tT_s - \tau_{P,m})^* = s_m(tT_s) - \hat{a}^{(I)}(tT_s - \hat{\tau}_{A,m}^{(I)}). \quad (17)$$

In Eq. 17, the recovered shifted P2 waveforms $p(tT_s - \tau_{P,m})^*$ are not obtained by simply considering $\hat{p}^{(I)}(tT_s - \tau_{P,m}^{(I)})$, but rather from the difference of the shifted A2 from the S2 signal. This choice has precedent in the literature, where existing approaches obtain the P2 via subtraction [4], [26].

III. NUMERICAL RESULTS WITH SYNTHETIC DATA

This set of experiments examines the performance of the proposed algorithm when compared to the unsupervised approach of [26], and the supervised approach of [27] for the separation of synthetic S2 sounds. The three evaluated methods share a similar set of assumptions regarding the variability of A2 and P2 components in different S2 sounds from a same heart sound recording, thus allowing a meaningful and fair comparison.

Subsection III-A evaluates how well the A2 and P2 waveforms are recovered as the difficulty of the separation task increases. Subsection III-B evaluates the robustness to

additive white Gaussian noise. Subsection III-C evaluates how well the predicted splitting delays correspond to the true splitting delays from the observed synthetic S2 signals.

Generation of Synthetic S2 signals: The numerical results shown in this section are obtained by considering synthetic signals generated using the model described in [18]. In particular, the S2 sounds are obtained as the superposition of A2 and P2 components modeled via Gaussian windowed sinusoids. The model is parameterized with respect to the frequency and phase of the sinusoidal signals as well as the amplitude, location, and width of the Gaussian window functions. These parameters are sampled from distributions obtained by fitting phonocardiogram (PCG) recordings obtained from a population of 150 subjects with age above 40 that were referred to a coronary computed tomography angiography with a low or intermediate pre-test probability for coronary artery disease (see [18] for details).

Proposed Model Hyperparameters: In all the reported results, the proposed alternating optimization method is run with a fixed number of iterations, $I = 10$, and the delay optimization step consists of 10 gradient descent iterations with step size $\gamma = 10^{-12}$ applied to frequency-domain signals computed over $N_{\text{DFT}} = 2048$ samples. The method in [27] is trained using 100 randomly generated heart sounds, each containing 40 heartbeats. The separation results for the three considered methods are averaged over 200 randomly generated heart sound recordings, each containing 40 heartbeats.

A. SEPARATION ACCURACY IN THE PRESENCE OF BIAS

A series of tests of increasing difficulty are conducted to evaluate how well the synthetic A2 and P2 waveforms are recovered from observed S2 signals. In the real world, some patients can have wider splitting delays between A2 and P2 components than others [6], [7]. When the splitting delay is small, the A2 and P2 components overlap and the separation problem is more difficult. To better understand how the models might be biased by real-world variability across patients, it is reasonable to test recovery when the splitting delay is smaller versus when it is larger. Secondly, within any given patient, inhalation typically increases the splitting delay and exhalation decreases the delay. We assume that, over a set of S2 sounds, the delays of S2 and P2 components, respectively, are uniformly distributed. To simultaneously capture statistics of inhalation within any given patient's S2 sound recording and expose bias to the natural variation of splitting found across different patients, a maximum delay $\Delta_{\text{max}} \in \{10, 15, 20, \dots, 80\}$ ms between the A2 and P2 components is introduced. For each maximum splitting delay Δ_{max} , a dataset of S2 sounds is generated by sampling splits from the uniform distribution $\mathcal{U}(0, \Delta_{\text{max}})$ ms. These tests are intended to capture splitting characteristics across patients, while assuming uniformly distributed splits within any patient.

The empirical measurement describing how well the A2 and P2 components are recovered is determined by

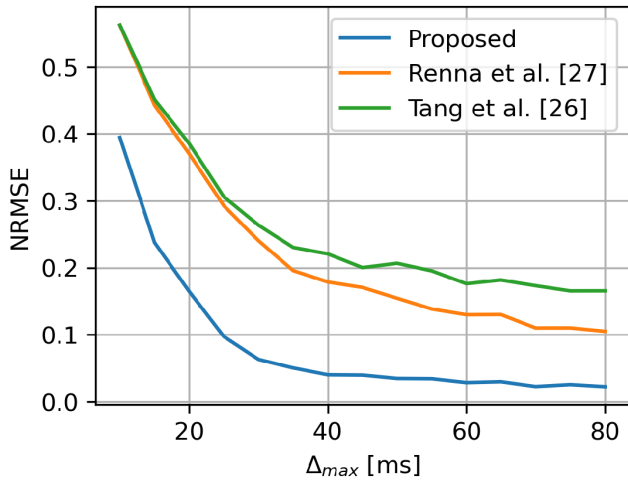


FIGURE 3. Recovers synthesized A2 and P2: The proposed algorithm is consistently best at recovery, for all tested Δ_{max} , including when the A2 and P2 components have small delay between them. The x-axis represents S2 sounds with wide or narrow splitting (i.e., large or small Δ_{max} , respectively) between A2 and P2 signals. All models favor wide splitting.

normalized root mean-squared error (NRMSE), defined as

$$\text{NRMSE} \triangleq \frac{\|a(tT_s - \tau_A) - a(tT_s - \tau_A^*)\|_2}{\|a(tT_s - \tau_A)\|_2} + \frac{\|p(tT_s - \tau_P) - p(tT_s - \tau_P^*)\|_2}{\|p(tT_s - \tau_P)\|_2}, \quad (18)$$

where $a(\cdot)$ and $p(\cdot)$ are the ground truth synthesized waveforms and $a(\cdot)^*$ and $p(\cdot)^*$ the predicted waveforms, and τ_A, τ_P, τ_A^* and τ_P^* are each vectors of length M containing the respective ground truth or predicted delay values. Note that τ_A^* is a zeros vector.

Fig. 3 visualizes the NRMSE values of the proposed separation algorithm and the methods in [26] and [27] under varying Δ_{max} . As expected, all methods show that NRMSE error increases as the separation task is made harder by greater overlap of the A2 and P2 components. The proposed approach recovers the synthetic A2 and P2 waveforms more precisely than the methods in [26] and [27] for all considered A2-P2 splits. Even when then the signals are largely overlapping in the time domain, where the separation problem is designed to be more difficult, the proposed method reconstructs waveforms with the lowest error (e.g. when $\Delta_{max} = 20$ ms, NRMSE is less than 0.2). The improvement over the related methods can be explained by observing that a key assumption in these baseline methods is relaxed in this work. Namely, for small values of Δ_{max} , the assumption that expected value of the P2 is approximately a flat signal

$$\frac{1}{M} \sum_m p(tT_s - \tau_{P,m}) \approx 0, \quad (19)$$

is too restrictive. For instance, Eq. (19) is explicitly defined in [26] in order for their method to obtain the A2 component by averaging the S2; our approach initializes A2 by

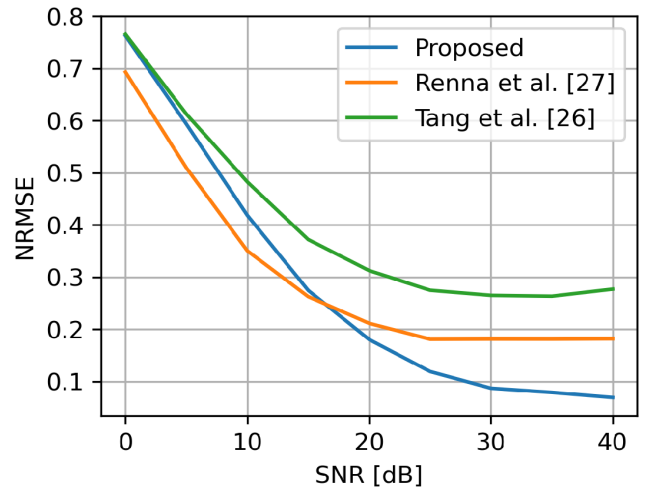


FIGURE 4. Robust to distortion by noise: The proposed method is better for SNR > 16 dB. Error increases as noise makes the signal increasingly random. The proposed unsupervised method can be outperformed by the supervised method [27] after significant noise corruption, and outperforms [26].

averaging, but then modifies it with alternating optimization. The delays obtained via the proposed approach increase flexibility and degrees of freedom of accurately recover the A2 and P2 waveforms.

B. ROBUSTNESS TO NOISE

Heart sound recordings are often subject to noise. A series of experiments test robustness to progressively increasing amounts of noise. In particular, when independent additive white Gaussian noise (AWGN) distorts the A2 and P2 components of the synthetically generated S2 sounds at different signal-to-noise ratio (SNR) levels, then the S2 sounds of different heartbeats are less well modeled by shifted versions of the same waveforms. Noise therefore challenges the key modeling assumption that the waveform is static across heartbeats. Fig. 4 reports the NRMSE obtained with the considered separation algorithms for different SNR values and for $\Delta_{max} = 40$ ms.

As shown in Fig. 4, the proposed approach outperforms the method in [26] for all considered SNR values greater than zero. The proposed approach does not require the A2 sounds must be perfectly aligned in all the observed S2 recordings, hence its superior performance. The proposed unsupervised method outperforms the supervised GMM-based separation method in [27] for levels of noise below approximately 16 dB, while the supervised approach gives empirically lower separation errors when there is significant noise (SNR values below 16 dB). Note that the considered GMM-based method requires access to separated A2 and P2 components for training, and this data is difficult or impossible to collect in real-world human heart sound data and the technology for practical application of the supervised method does not yet exist. With similar reasoning, continuing research on noise cancellation approaches would improve the observed SNR,

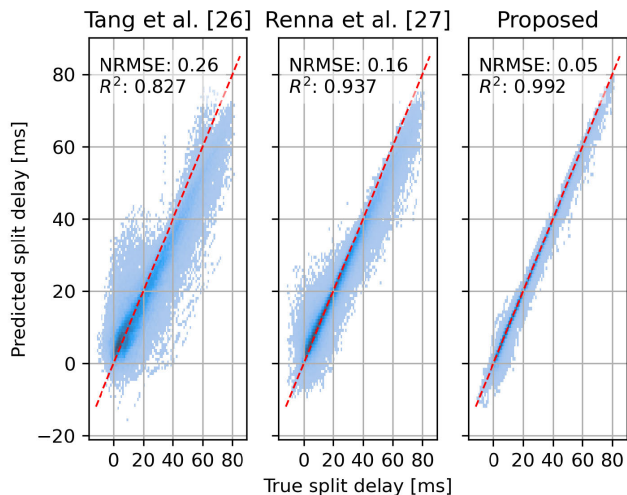


FIGURE 5. Superior recovery of split delays between the synthesized A2 and P2 components. Compared to approaches [26] (left) and [27] (middle), the proposed method (right) has largest R² correlation and lowest error. Each plot visualizes a scatter plot of 120k samples. Red line is ideal.

and this serves to benefit the proposed algorithm over other methods.

C. RECOVERY OF A2-P2 SPLIT DELAYS

Time delays between the A2 and P2 components of an S2 sound are clinically relevant, and estimation of delays should be accurate and minimally biased. We compare recovered time delays Δ_m^{*} produced by the considered separation algorithms to corresponding ground truth time delays Δ_m on a dataset of M = 120000 synthetic heart sounds.

All delays are computed via the same procedure. Given as input the A2 and P2 components of either a ground truth signal or recovered signal, a delay is computed as the time difference between the main peak of the P2 component and that of the A2 component. For each component, its main peak is computed using the homomorphic envelopgram [29] of the component [26]. A homomorphic envelopgram of a signal is a strictly positive representation of the signal that smooths out splits and serrated peaks in the signal, and it is obtained by applying a low pass filter to the natural logarithm of the component and then exponentiating. For instance, the homomorphic envelopgram of the P2 component is h_{P,m}(tT_s) = exp(L(ln(p(tT_s - τ_{P,m})))) for a low pass filter L(·). In Eq. (20), the main peak is obtained as the weighted sum of the time instants, with weights given by the normalized homomorphic envelopgram. Denoting h_{A,m}(tT_s) and h_{P,m}(tT_s) for the homomorphic envelopgrams of the A2 and P2 components of the m-th recorded S2 sounds, the ground truth A2-P2 split is computed as

$$\Delta_m = \frac{\sum_t h_{P,m}(tT_s) \cdot tT_s}{\sum_t h_{P,m}(tT_s)} - \frac{\sum_t h_{A,m}(tT_s) \cdot tT_s}{\sum_t h_{A,m}(tT_s)}. \quad (20)$$



FIGURE 6. Digiscope Collector technology used to record the heart sounds [33] in the considered real-world dataset.

Given a vector of M ground truth delays Δ, and M predicted delays Δ^{*}, the normalized mean squared error is

$$\text{NRMSE}_\Delta \triangleq \frac{\|\Delta - \Delta^*\|_2}{\|\Delta\|_2}. \quad (21)$$

The scatter plots in Fig. 5 compare the ground truth delays of the A2-P2 splits as obtained from the generative signal model to the corresponding estimates obtained from the source separation algorithms. The proposed separation method has the lowest error. It has an NRMSE_Δ of 0.05 compared to 0.16 for the algorithm in [27] and 0.26 for that of [26]. It also has the lowest R² coefficient of determination.

IV. APPLICATION TO REAL-WORLD DATA

We also tested the proposed separation algorithm on real-world data, in order to assess its potential in a real, clinical environment. In this scenario, in the presence of only PCG recordings, ground truth values for the waveforms associated with the A2 and P2 components of S2 sounds are not directly available. In the absence of separated components for training, the supervised GMM-based separation algorithm of [27] cannot be considered for comparison in this section. Moreover, the separation performance of the proposed method is gauged indirectly by evaluating the statistical significance for detecting PH of features extracted from the resulting separated A2 and P2 components.

The features analyzed are chosen to reflect the observation that specific characteristics of the S2 have been linked to the pulmonary blood pressure. In particular, features extracted from separated A2 and P2 components, as the time split, peak-to-peak ratio, energy ratio, etc., are widely recognized as potential indicators of PH [5], [30], [31]. In fact, the analysis of the S2 has been studied as a potential way to provide non-invasive estimates of the PAP [4], [32].

A. MATERIALS

The heart sounds used for this study were collected in the Real Hospital Português, Brazil, using a Littmann 3200 electronic stethoscope, and the DigiScope Collector software [33] (Fig. 6). Auscultation was performed over the second left

intercostal space, collecting sounds of duration approximately 60 seconds in a quiet environment. All collected heart sounds were collected at 4 kHz and the amplitudes were quantized to 8-bit resolution. The heart sounds were anonymized and analyzed in Portugal with the approval of Real Hospital Português and the Ethics Committees of University of Porto, Portugal (approval reference number 391.391 of September 6th, 2013).

The considered dataset comprises a total of 43 heart sound recordings collected from 43 adult patients, 11 of which were diagnosed with PH, while the remaining 32 were not diagnosed with PH.

B. PRE-PROCESSING

The raw PCG signals were first filtered with high-pass and low-pass Butterworth filters of order two with cut-off frequencies equal to 25 Hz and 400 Hz, respectively, in order to keep the main spectral information associated to both A2 and P2 components while rejecting out-of-band noise. Then, the spike removal algorithm presented in [34] was applied to the filtered signals to address ambient noise, and then each recording was normalized to zero mean and unit variance.

C. SEGMENTATION

The pre-processed PCG signals were segmented to determine the exact localization of S2 sounds using the combined convolutional neural network (CNN) and hidden Markov model (HMM) approach presented in [35]. However, further manual correction was applied, due to the presence of segmentation errors induced by variable heart rates and due to the presence of significant systolic murmurs in some of the patients affected by PH.

D. S2 ALIGNMENT

The S2 segments extracted from each PCG recording are first aligned, in order to allow the use of the initialization procedure described in Section II-B and in order to facilitate the extraction of the A2 and P2 components. Alignment is performed in two stages: (i) coarse alignment and (ii) fine alignment.

Coarse alignment is obtained by computing first the homomorphic envelopogram [36] of the PCG to determine the time intervals of S2 sounds where the majority of the energy is contained. The homomorphic envelopograms of the different S2 sounds extracted from a given PCG are aligned using a cross-correlation analysis. In particular, for each of the S2 sounds, the alignment shift was obtained as the lag index that maximizes the correlation between the first S2 sound of the recording and the given S2 sound within a search window of 200 ms centered around the zero-lag index.

Then, fine alignment is obtained by applying the approach again, and directly to the filtered PCG signal of the coarsely aligned S2 segments. Also in this case, the lag index that maximizes the cross-correlation between the first S2 in the

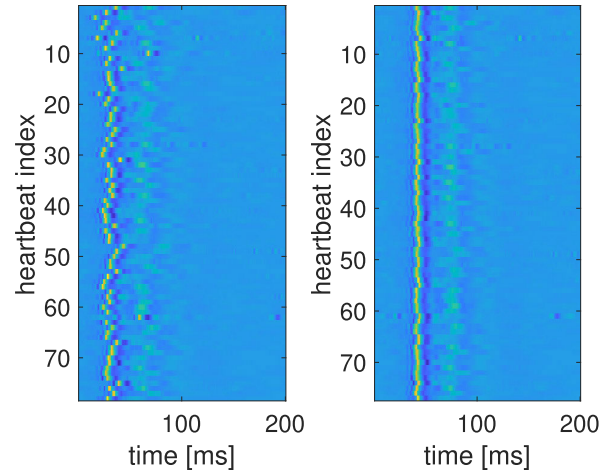


FIGURE 7. Example S2 sounds extracted from a PCG recording of the considered real-world dataset before (left) and after (right) the alignment procedure. Periodicity of respiratory phase is visible in the aligned S2 sounds.

recording and the given S2 sound is searched for within a window of 200 ms centered around the zero-lag index. Moreover, only segments corresponding to a normalized correlation coefficient greater than 0.2 are retained, in order to exclude S2 segments affected by excessive artifacts and noise levels.

Note that the proposed two-step alignment procedure mainly aims to increase robustness against noise by considering the homomorphic envelopogram of the S2 sounds in the first step. Fig. 7 reports an example of S2 sounds extracted from a PCG recording of the considered dataset before and after the two-steps alignment procedure. Fig. 1 shows an example of the source separation into the A2 and P2 components.

E. DISCRIMINATING PH

First, the separation of A2 and P2 components is performed by applying the procedure described in Section II or the method described in [26] to the aligned S2 segments of a PCG recording.

Then, features are extracted from the obtained A2 and P2 components in order to discriminate heart sounds recorded from patients with and without PH. In particular, it has been observed that, most commonly, heart sounds of patients with PH present louder P2 components [31], along with larger A2-P2 time splits [4].

The features extracted from the recorded PCG signals are also extracted from the outputs of our proposed separation algorithm, and also the algorithm of [26]. They are:

- *Mean time split:* We compute the mean relative time delay between A2 and P2 components from the homomorphic envelope as it was defined in Eq. 20,

$$M_{\hat{\Delta}} = \frac{1}{M} \sum_{m=0}^{M-1} \hat{\Delta}_m. \quad (22)$$

TABLE 1. Meaningful features for PH detection on real world data are obtained via our proposed method. The first three rows of the table verify that the features extracted from proposed A2 and P2 components of the S2 sound are each independently discriminative of PH. The values in these first three rows are p -values corresponding to two sample Kolmogorov-Smirnov tests of PH and not-PH patients. The remaining rows report predictive performance of a PH classifier trained jointly on the three features. Columns compare the algorithm presented in [26] with the proposed method.

	Tang <i>et al.</i> [26]	Proposed
M_{Δ}	0.0013	0.0056
R_E	0.0108	0.0071
R_P	0.0007	0.0013
auROC	0.72	0.76
AP	0.38	0.41

- *Energy ratio*: This feature is obtained by computing the average over the S2 sounds contained in a PCG of the ratio between the energy of the P2 component and the energy of the corresponding S2 sounds, i.e.,

$$R_E = \frac{1}{M} \sum_{m=0}^{M-1} \frac{\sum_t |p_m(tT_s)|^2}{\sum_t |s_m(tT_s)|^2}. \quad (23)$$

- *Peak-to-peak ratio*: This feature is obtained by computing the average over the S2 sounds contained in a PCG of the ratio between the peak-to-peak amplitude of the P2 component and of the corresponding S2 sound, i.e.,

$$R_P = \frac{1}{M} \sum_{m=0}^{M-1} \frac{|\max_t \{p_m(tT_s)\} - \min_t \{p_m(tT_s)\}|}{|\max_t \{s_m(tT_s)\} - \min_t \{s_m(tT_s)\}|}. \quad (24)$$

Table 1 demonstrates that features derived from the proposed A2 and P2 components are predictive of PH. The first three rows verify that each of the three features independently are useful for PH via two sample Kolmogorov-Smirnov (KS) statistical significance tests. The KS tests evaluate the null hypothesis that there is no difference, given the feature and dataset, of whether a subject has PH. The KS test supports varying size samples; there are 11 PH negative and 32 PH positive samples in each significance test. The remaining rows evaluate all three features jointly with a predictive model trained to detect PH. The classifier model is a kernel ridge regression classifier with two degree polynomial kernel trained on the three features described in the table, and it was evaluated via leave one out cross validation with metrics including the area under the ROC curve (auROC) and Average Precision (AP). For comparison to the method in [26], the first two steps in Fig. 2 and the features obtained are identical for both methods, and only the obtained A2 and P2 components depend on the method. The results suggest that the proposed model is at least as accurate as the method of [26]. The features analyzed verify that the A2 and P2 components obtained by our proposed method contain information useful for discrimination of PH.

V. DISCUSSION

The proposed separation method stems from the observation that heart sounds are modulated by the respiratory activity.

This observation has also been made in previous state-of-the-art approaches in [26] and [27]. In particular, inspiration increases displacement between the A2 and P2 components while exhalation decreases displacement. This observation implies that averaging aligned and segmented S2 signals in order to reduce noise may introduce significant distortion to P2 components. In fact, the method in [26] relies on simple averaging to completely remove P2 components from S2 sounds, thus obtaining A2 estimates.

The proposed approach enhances this formulation with a more flexible mathematical framework to recover the A2 and P2 components. It does not require any prior information on the specific shape of such waveforms, and it does not impose the assumption that averaging removes the P2 components. The proposed method is shown to have accurate and robust separation performance, including when aortic and pulmonary components have small delay between them.

The computational complexity of the proposed alternating optimization method is mainly dominated by a pseudoinverse, which has computational complexity of $\mathcal{O}(MN^3)^2$ and a frequency-domain projected gradient descent with complexity $\mathcal{O}(MN \log N)$.

Future work: The proposed method achieves automatic separation of previously segmented S2 sounds. Future work should automate the pre-processing steps to denoise, segment and align the S2 signals. A future adaptation of this work could improve robustness to severe noise or could consider an objective that relaxes the assumption during optimization that the waveform generating function is identical across heartbeats.

VI. CONCLUSION

We have presented a novel unsupervised blind source separation algorithm to identify the A2 and P2 components and a set of delays representative of respiration rate from S2 recordings. The proposed approach has the advantage of flexibly modeling aortic and pulmonary components without using predefined waveforms. The reconstruction of the A2 and P2 components is defined in terms of an optimization problem whose solution is approximated via alternating optimization of a least squares objective.

The proposed separation method is shown to provide accurate and robust reconstructions of the aortic and pulmonary components, and the method can outperform previous approaches in the literature that were formulated on similar premises. We report consistent improvement of 10 percent lower normalized root mean-squared error (NRMSE) over the state-of-the-art in the reconstruction of aortic and pulmonary components on a synthetic dataset. Moreover, the proposed approach has higher coefficient of correlation (R^2) and lower NRMSE than comparable related works when recovering delays between components on a synthetic dataset. The proposed method allows for

²Note that, given the sparsity patterns characterizing the matrices $\mathbf{X}^{(i)}$, efficient inversion algorithms specifically designed for sparse matrices could be used to find the solution of the corresponding linear least-squares problem.

the extraction of discriminative features from A2 and P2 components for detection of pulmonary hypertension from real-world heart sounds, and on the evaluated dataset, the features were shown give superior classification performance to a comparable competing method.

The proposed approach needs only segmented S2 heart sounds; it does not require a respiration rate measurement, annotations of the A2 and P2 components, or other auxiliary data. The presented separation method establishes the foundation for a series of improvements that will be the focus of future research.

REFERENCES

- [1] W. H. Organization. *Fact Sheet: Cardiovascular Diseases (CVDs)*. World Health Organization. Accessed: Oct. 25, 2022. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs317/en/>
- [2] G. Butrous, H. A. Ghofrani, and F. Grimminger, "Pulmonary vascular disease in the developing world," *Circulation*, vol. 118, no. 17, pp. 1758–1766, Oct. 2008.
- [3] C. E. Angermann, B. Assmus, S. D. Anker, F. W. Asselbergs, J. Brachmann, M.-E. Brett, J. J. Brugs, G. Ertl, G. Ginn, and L. Hilker, "Pulmonary artery pressure-guided therapy in ambulatory patients with symptomatic heart failure: The CardioMEMS European monitoring study for heart failure (MEMS-HF)," *Eur. J. heart failure*, vol. 22, no. 10, pp. 1891–1901, 2020.
- [4] J. Xu, "A new, simple, and accurate method for non-invasive estimation of pulmonary arterial pressure," *Heart*, vol. 88, no. 1, pp. 76–80, Jul. 2002.
- [5] R. Smith and D. Ventura, "A general model for continuous noninvasive pulmonary artery pressure estimation," *Comput. Biol. Med.*, vol. 43, no. 7, pp. 904–913, Aug. 2013.
- [6] R. F. Castle and K. L. Jones, "The mechanism of respiratory variation in splitting of the second heart sound," *Circulation*, vol. 24, no. 2, pp. 180–184, Aug. 1961.
- [7] G. Amit, K. Shukha, N. Gavrieli, and N. Intrator, "Respiratory modulation of heart sound morphology," *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 296, no. 3, pp. H796–H805, Mar. 2009.
- [8] R. Guedes, H. C. Carvalho, and A. Castro, "Heart sound analysis for blood pressure estimation," in *Advanced Methodologies and Technologies in Medicine and Healthcare*. Hershey, PA, USA: IGI Global, 2019, pp. 98–109.
- [9] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, 1st ed. Orlando, FL, USA: Academic, 2010.
- [10] J. Bobin, J.-L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2662–2674, Nov. 2007.
- [11] V. Nigam and R. Priemer, "A dynamic method to estimate the time split between the A2 and P2 components of the S2 heart sound," *Physiol. Meas.*, vol. 27, no. 7, pp. 553–567, Jul. 2006.
- [12] Z. Tong, I. A. Qader, and F. Abu-Amara, "Heart sound separation using fast independent component analysis," in *Proc. Int. Conf. Develop. E-Systems Eng. (DeSE)*, Dec. 2015, pp. 3–6.
- [13] M. Mørup, K. H. Madsen, and L. K. Hansen, "Shifted independent component analysis," in *Proc. Int. Conf. Independ. Compon. Anal. Signal Separat.* Cham, Switzerland: Springer, 2007, pp. 89–96.
- [14] J. Xu, L. Durand, and P. Pibarot, "Nonlinear transient chirp signal modeling of the aortic and pulmonary components of the second heart sound," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 10, pp. 1328–1335, Oct. 2000.
- [15] B. Popov, G. Sierra, L.-G. Durand, J. Xu, P. Pibarot, R. Agarwal, and V. Lanzo, "Automated extraction of aortic and pulmonary components of the second heart sound for the estimation of pulmonary artery pressure," in *Proc. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 1, Sep. 2004, pp. 921–924.
- [16] R. Agarwal, E. I. Plotkin, and M. N. S. Swamy, "Statistically optimal null filter based on instantaneous matched processing," *Circuits Syst. Signal Process.*, vol. 20, no. 1, pp. 37–61, Jan. 2001.
- [17] A. Djebbari and F. Berekci-Reguig, "A new chirp-based wavelet for heart sounds time–frequency analysis," *Int. J. Commun. Antenna Propag. (IRECAP)*, vol. 1, pp. 92–102, 2011.
- [18] R. G. Sæderup, P. Hoang, S. Winther, M. Böttcher, J. Struijk, S. Schmidt, and J. Østergaard, "Estimation of the second heart sound split using windowed sinusoidal models," *Biomed. Signal Process. Control*, vol. 44, pp. 229–236, Jul. 2018.
- [19] T. S. Leung, A. P. Salmon, W. B. Collis, P. R. White, E. Brown, and J. Cook, "Analysis of the second heart sound for diagnosis of paediatric heart disease," *IEE Proc. Sci., Meas. Technol.*, vol. 145, no. 6, pp. 285–290, Nov. 1998.
- [20] J. Xu, L.-G. Durand, and P. Pibarot, "Extraction of the aortic and pulmonary components of the second heart sound using a nonlinear transient chirp signal model," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 3, pp. 277–283, Mar. 2001.
- [21] S. M. Debbal and F. Berekci-Reguig, "Automatic measure of the split in the second cardiac sound by using the wavelet transform technique," *Comput. Biol. Med.*, vol. 37, no. 3, pp. 269–276, Mar. 2007.
- [22] S. R. Thiagaraja, J. Vempati, R. Dantu, T. Sarma, and S. Dantu, "Smart phone monitoring of second heart sound split," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 2181–2184.
- [23] S. Barma, B.-W. Chen, K. L. Man, and J.-F. Wang, "Quantitative measurement of split of the second heart sound (S2)," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 12, no. 4, pp. 851–860, Jul. 2015.
- [24] S. Santos, R. Paiva, P. Carvalho, and J. Henriques, "Detection of the S2 split using the Hilbert and wavelet transforms," in *Congresso de Métodos Numéricos em Engenharia*, Coimbra, Portugal, 2011, pp. 1–11.
- [25] D. Chen, P. Pibarot, G. Honos, and L.-G. Durand, "Estimation of pulmonary artery pressure by spectral analysis of the second heart sound," *Amer. J. Cardiol.*, vol. 78, no. 7, pp. 785–789, Oct. 1996.
- [26] H. Tang, H. Chen, and T. Li, "Discrimination of aortic and pulmonary components from the second heart sound using respiratory modulation and measurement of respiratory split," *Appl. Sci.*, vol. 7, no. 7, p. 690, Jul. 2017.
- [27] F. Renna and M. Coimbra, "Source separation of the second heart sound using Gaussian mixture models," in *Proc. Comput. Cardiol. (CinC)*, Sep. 2019, pp. 1–4.
- [28] F. Renna, M. D. Plumbley, and M. Coimbra, "Source separation of the second heart sound via alternating optimization," in *Proc. Comput. Cardiol. (CinC)*, vol. 48, Sep. 2021, pp. 1–4.
- [29] D. Gill, N. Gavrieli, and N. Intrator, "Detection and identification of heart sounds using homomorphic envelopment and self-organizing probabilistic model," in *Proc. Comput. Cardiol.*, Sep. 2005, pp. 957–960.
- [30] M. Elgendy, P. Bobhate, S. Jain, J. Rutledge, J. Y. Coe, R. Zemp, D. Schuurmans, and I. Adatia, "Time-domain analysis of heart sound intensity in children with and without pulmonary artery hypertension: A pilot study using a digital stethoscope," *Pulmonary Circulat.*, vol. 4, no. 4, pp. 685–695, Dec. 2014.
- [31] L. Bickley and P. G. Szilagy, *Bates' Guide to Physical Examination and History-Taking*. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2012.
- [32] T. Kaddoura, K. Vadlamudi, S. Kumar, P. Bobhate, L. Guo, S. Jain, M. Elgendy, J. Y. Coe, D. Kim, D. Taylor, W. Tymchak, D. Schuurmans, R. J. Zemp, and I. Adatia, "Acoustic diagnosis of pulmonary hypertension: Automated speech-recognition-inspired classification algorithm outperforms physicians," *Sci. Rep.*, vol. 6, no. 1, pp. 1–11, Sep. 2016.
- [33] P. Gomes, S. Frade, A. Castro, R. Cruz-Correia, and M. Coimbra, "A proposal to incorporate digital auscultation and its processing into an existing electronic health record," in *Proc. Int. Conf. Health Informat.*, 2015, pp. 143–150.
- [34] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of heart sound recordings by a duration-dependent hidden Markov model," *Physiol. Meas.*, vol. 31, no. 4, pp. 513–529, Apr. 2010.
- [35] F. Renna, J. Oliveira, and M. T. Coimbra, "Deep convolutional neural networks for heart sound segmentation," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 6, pp. 2435–2445, Nov. 2019.
- [36] I. Rezek and S. J. Roberts, "Envelope extraction via complex homomorphic filtering," Imperial College London, London, U.K., Tech. Rep. TR-98-9, 1998.



FRANCESCO RENNA (Senior Member, IEEE) received the Laurea Specialistica degree in telecommunication engineering and the Ph.D. degree in information engineering from the University of Padova, Padua, Italy, in 2006 and 2011, respectively. Between 2007 and 2022, he was a Visiting Researcher and a Postdoctoral Researcher with Infineon Technology AG, Princeton University, Georgia Institute of Technology (Lorraine Campus), Supélec, University of Porto, Duke University, University College London, and University of Cambridge. Since 2022, he has been a Senior Researcher with INESC TEC and an Assistant Professor with the Department of Computer Science, Faculty of Sciences, University of Porto. His research interests include high-dimensional information processing for biomedical signal and image analysis. He was a recipient of a Marie Skłodowska-Curie Individual Fellowship from the European Commission and a Research Contract within the Scientific Employment Stimulus Program, Portuguese Foundation for Science and Technology.



MARK D. PLUMBLEY (Fellow, IEEE) received the Ph.D. degree in neural networks from the University of Cambridge, Cambridge, U.K., in 1991. He is currently a Professor in signal processing with the Centre for Vision, Speech, and Signal Processing and is the Head of the School of Computer Science and Electronic Engineering, University of Surrey, Guildford, U.K. He is also an Expert on the analysis and processing of audio, using a wide range of signal processing and machine learning methods. He led the first international data challenge on detection and classification of acoustic scenes and events and currently holds an engineering and physical sciences research Council Fellowship for “AI for Sound,” on the automatic recognition of everyday sounds.



ALEX GAUDIO (Member, IEEE) received the B.A. degree in jazz performance and composition from the Bard College, the M.S. and Ph.D. degrees from Carnegie Mellon University (CMU), and the Ph.D. degree from the University of Porto, in 2023. He was honored as a Champion of Social Justice for his non-profit, NYC Makerspace, in 2018. He was a Data Scientist and an Engineer in New York City for seven years (2011–2018). He was also an Emergency Medical Technician in ambulances and in a hospital during the B.A. degree. His researches explainable machine learning and signal processing for the analysis of heart and lung sounds.



MIGUEL TAVARES COIMBRA (Senior Member, IEEE) was one of the founders of IT Porto, in 2007, it's a Coordinator, from 2015 to 2019, and the Founder of the Interactive Media Group at this institute. He was the Director of the Master in Medical Informatics with the University of Porto, from 2014 to 2016, and was the Co-Founder of IS4H—Interactive Systems for Healthcare, a spin-off company of the University of Porto, in 2013, where he has been a member of the Executive Board of the Faculty of Sciences, since 2019, the current Coordinator of the TEC4Health Line of INESC TEC, and the past Chair of the Portugal Chapter of the IEEE Engineering and Medicine Society (2017–2021). He is currently a Full Professor with the Computer Science Department, Faculty of Sciences, University of Porto. He leads and participates in various projects involving engineering and medicine, namely cardiology and gastroenterology, with current and past collaborations with hospitals in Portugal, Brazil (Pernambuco, Paraíba, Minas Gerais, and São Paulo), Germany, and Sweden. The nearly 16 years of experience in biomedical signal processing and interactive systems for healthcare have led to the development and deployment of systems for the collection and analysis of auscultation signals, echocardiography image processing for rheumatic fever screening, monitoring of stress and fatigue of firefighters in action, endoscopy signal analysis for cancer detection, computer-assisted decision systems for capsule endoscopy, and quantification of 3D motion patterns for epilepsy. He has more than 130 scientific publications, 25 of which in high-impact scientific journals (17 IEEE TRANSACTIONS) and has attracted and managed more than 2M€ in research funding, over a total of 15 research projects acting as a PI of the project (ten projects) or a co-PI of its institution (five projects).



SANDRA MATTOS received the joint Medical degree in pediatric and fetal cardiology from the Federal University of Pernambuco, Royal Brompton Hospital, and Great Ormond Street Hospital, London, U.K., in 1981, and the Ph.D. degree in northeast biotechnology network/RENORBIO from the Keiko Asami Immunopathology Laboratory, Federal University of Pernambuco. Her Ph.D. thesis in development with the Keiko Asami Immunopathology Laboratory, Federal University of Pernambuco. She is formerly a Locumn Consultant in paediatric cardiology with Radcliffe Hospital, Oxford, U.K., and an Assistant Professor in peds card with the University of North Carolina, Chapel Hill, USA. Currently, she is the Director of the Maternal-Fetal Cardiac Unit and the Caduceus-Research Center, Royal Portuguese Hospital, Recife, and a Coordinator of the Heart Circle of Pernambuco, Brazil. Her expertise areas are pediatric and fetal cardiology and echocardiography.

...