Source Separation of the Second Heart Sound Using Gaussian Mixture Models

Francesco Renna[1], Miguel Coimbra[2]

[1] Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto, Portugal
[2] INESC TEC, Faculdade de Ciências da Universidade do Porto, Portugal

Abstract

In this work, we present a method to separate aortic (A2) and pulmonary (P2) components from second heart sounds (S2). The proposed approach captures the different dynamical behavior of A2 and P2 components via a joint Gaussian mixture model, which is then used to perform separation via a closed-form conditional mean estimator.

The proposed approach is tested over synthetic heart sounds and it is shown guarantee a reduction of approximately 25% of the normalized root mean-squared error incurred in signal separation, with respect to a previously presented approach in the literature.

1. Introduction

Cardiac auscultation represents arguably the most cost-effective screening method for a number of heart diseases. Precious diagnostic information, especially regarding the mechanical activity of heart valves, can be retrieved from the analysis of heart sounds. In addition to diseases signalled by the presence of extra sounds like murmurs, clicks, etc., it is possible to extract useful information from the analysis of the waveforms associated to fundamental heart sounds. In particular, the second heart sound (S2) is formed by two constituent components: one generated by the closure of the aortic valve (usually called the A2 component) and one generated by the closure of the pulmonary valve (which is called the P2 component). The analysis of the P2 component has been attracting attention as a possible method to estimate the pulmonary artery pressure (PAP) in a noninvasive manner, thus avoiding right heart catheterization [1, 2].

On the other hand, the separation of the A2 and P2 components of S2 sounds represents a very challenging problem, due to their large overlap in the time and frequency domain and their morphological similarity. Moreover, A2 and P2 components cannot be modeled as independent signals, due to the interaction of the mechanical processes that originate them. For these reasons, the application of standard methods for blind source separation as independent component analysis (ICA) [3] and morphological compo-

nent analysis (MCA) [4] are not suitable to solve the A2-P2 separation problem.

Therefore, some separation approaches specifically tailored to the characteristics of S2 signals have appeared in the literature. Some of these approaches assume that A2 and P2 components can be modeled accurately via a pre-determined kind of waveform (e.g., chirp signals [1], windowed sinusoids [5], etc.) whose parameters are then fitted to the recorded S2 data via different optimization methods. A different approach has recently emerged which refrains from adopting a predetermined waveform model and instead uses the different dynamical behavior of A2 and P2 components in different S2 sounds to perform separation. In particular, Tang *et al.* [6] leverage the fact that A2 components usually occupy the same position in different S2 sounds, whereas P2 components experience a shift according to the respiration phase of each recorded S2 sound (i.e., P2 components experience a more significant delay from A2 components during inspiration phases than during expiration phases). Therefore, assuming that A2 and P2 components keep approximately the same form (apart from time shifts) in successive S2 sounds, the A2 components are estimated by averaging over the collected S2 sounds and the P2 components are obtained by subtraction.

Building on similar modeling assumptions, the proposed approach aims to capture the details of A2 and P2 components by using a joint Gaussian mixture model (GMM) prior. In particular, the proposed approach learns the shape of A2 and P2 from training data and uses such knowledge to separate the constituents of unseen S2 sounds. The choice of GMM priors is determined by their ability in representing different kinds of real-world signals, as patches extracted from natural images [7], heart sound segments [8], etc., and their mathematical amenability in deriving fundamental limits and closed-form algorithms for signal recovery [7, 9, 10]. Therefore, inspired by recent results on the use of a GMM to perform signal de-mixing of [10], the proposed method is based on the following steps:

1. Collection and alignment of different S2 sounds from a given heart sound recording.
2. Modeling of patches extracted from multiple A2 and P2 sounds via a joint GMM.

3. Separation of unseen S2 patches into their components using a closed-form conditional mean estimator.

Then, the potential of the proposed approach is tested over a set of synthetic signals generated according to the windowed sinusoid model presented in [5].

The following notation is adopted in the remainder of this paper. The identity matrix of dimension $n \times n$ is denoted by $\mathbf{I_n}$. The transpose and Moore-Penrose pseudoinverse operators are denoted by $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\dagger}$, respectively. The multivariate Gaussian distribution at a vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbb{E}[\cdot]$ denotes the expectation operator.

2. Methods

In the following, we present the mathematical model used to describe A2 and P2 components and the corresponding source separation algorithm.

2.1. Joint GMM for A2 and P2 components

We assume that heart sounds are acquired with a given sampling frequency F_s, filtered, normalized, and then segmented in order to determine the position and the boundaries of the corresponding S2 sounds. The samples associated to the i-th S2 sounds extracted from a given recording are denoted by $s_i(n)$ for $n = 0, \ldots, N-1$ and $i = 0, \ldots, I-1$, and we assume that

$$s_i(n) = a_i(n) + p_i(n), \qquad (1)$$

where $a_i(n)$ and $p_i(n)$ represent the A2 and P2 components, respectively.

The proposed separation method is based on the observation of consecutive S2 sounds from a given heart sound recording. In particular, sample vectors of interest are obtained by picking N_t samples from N_b consecutive S2 sounds and by stacking them in a single $N_t N_b \times 1$ vector as follows:

$$\mathbf{s_{m,j}} = \left[\bar{\mathbf{s}}_{\mathbf{m,j}}^{\mathrm{T}}, \bar{\mathbf{s}}_{\mathbf{m,j+1}}^{\mathrm{T}}, \ldots, \bar{\mathbf{s}}_{\mathbf{m,j+N_b-1}}^{\mathrm{T}} \right]^{\mathrm{T}}, \qquad (2)$$

where $\bar{\mathbf{s}}_{m,j} = [s_j(m), s_j(m+1), \ldots, s_j(m+N_t-1)]^{\mathrm{T}}$. This means that the vector $\mathbf{s}_{m,j}$ contains N_t samples starting from the index m from N_b consecutive S2 sounds in a heart sound recording, starting from the heartbeat with index j. A similar construction can be applied to the A2 and P2 components in (1), thus obtaining the vectors $\mathbf{a_{m,j}}$ and $\mathbf{p_{m,j}}$, respectively, such that $\mathbf{s_{m,j}} = \mathbf{a_{m,j}} + \mathbf{p_{m,j}}$.

Then, on defining the vector $\mathbf{x_{m,j}} = [\mathbf{a_{m,j}^{\mathrm{T}}}, \mathbf{p_{m,j}^{\mathrm{T}}}]^{\mathrm{T}}$, we can also use the following matrix notation:

$$\mathbf{s_{m,j}} = \boldsymbol{\Phi} \mathbf{x_{m,j}}, \qquad (3)$$

where

$$\boldsymbol{\Phi} = [\mathbf{I_{N_t \cdot N_b}}, \mathbf{I_{N_t \cdot N_b}}]. \qquad (4)$$

Irrespective of the indexes m and j, the vectors $\mathbf{a_{m,j}}$ and $\mathbf{p_{m,j}}$ containing sample from A2 and P2 components from different S2 sounds are modeled via a joint GMM. Therefore, in order to simplify that notation, we can drop the subscripts and assume that vectors $\mathbf{x} = [\mathbf{a}^{\mathrm{T}}, \mathbf{p}^{\mathrm{T}}]^{\mathrm{T}}$ are drawn from the multivariate distribution

$$p(\mathbf{x}) = \sum_{\mathbf{k=1}}^{\mathbf{K}} \pi_{\mathbf{k}} \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{\mathbf{x}}^{(\mathbf{k})}, \boldsymbol{\Sigma}_{\mathbf{x}}^{(\mathbf{k})}), \qquad (5)$$

where K is the number of Gaussian components in the GMM, and the k-th Gaussian component has mean

$$\boldsymbol{\mu}_{\mathbf{x}}^{(k)} = \left[\begin{array}{c} \boldsymbol{\mu}_{\mathbf{a}}^{(k)} \\ \boldsymbol{\mu}_{\mathbf{p}}^{(k)} \end{array} \right] \qquad (6)$$

and covariance matrix

$$\boldsymbol{\Sigma}_{\mathbf{x}}^{(\mathbf{k})} = \left[\begin{array}{cc} \boldsymbol{\Sigma}_{\mathbf{a}}^{(\mathbf{k})} & \boldsymbol{\Sigma}_{\mathbf{ap}}^{(\mathbf{k})} \\ \boldsymbol{\Sigma}_{\mathbf{pa}}^{(\mathbf{k})} & \boldsymbol{\Sigma}_{\mathbf{p}}^{(\mathbf{k})} \end{array} \right]. \qquad (7)$$

2.2. Separation algorithm

The proposed separation algorithm is based on modeling portions of A2 and P2 components via a joint GMM and it consists of two phases: training and separation. In the training phase, samples from available A2 and P2 components are collected to form a training dataset of vectors $\mathbf{x}$. Then, using such data, the parameters characterizing the joint GMM distribution in (5), i.e., the component probabilities π_k, the mean vectors $\boldsymbol{\mu}_{\mathbf{x}}^{(k)}$ and the covariance matrices $\boldsymbol{\Sigma}_{\mathbf{x}}^{(\mathbf{k})}$, are estimated using the expectation maximization (EM) algorithm [11].

Then, the separation phase consists in recovering vectors $\mathbf{x}$ from the observation of the corresponding vectors $\mathbf{s} = \boldsymbol{\Phi} \mathbf{x}$. Such inversion can be performed via the conditional mean estimator

$$\hat{\mathbf{x}} = \mathbb{E}[\mathbf{x}|\mathbf{s}] = \int \mathbf{x} \, p(\mathbf{x}|\mathbf{s}) d\mathbf{x}, \qquad (8)$$

which can be shown to minimize the mean-squared error $\mathrm{MSE} = \mathbb{E}\left[|\mathbf{x} - \hat{\mathbf{x}}|^{\mathbf{2}} \right]$. Most notably, modeling signals of interest with joint GMMs implies that the corresponding conditional mean estimator can be expressed in closed form. Therefore, the adopted separation algorithm can be expressed as [10]:

$$\hat{\mathbf{x}} = \sum_{k=1}^{K} \tilde{\pi}_k \left(\boldsymbol{\mu}_{\mathbf{x}}^{(k)} + \boldsymbol{\Sigma}_{\mathbf{x}}^{(\mathbf{k})} \boldsymbol{\Phi}^{\mathrm{T}} (\boldsymbol{\Phi} \boldsymbol{\Sigma}_{\mathbf{x}}^{(\mathbf{k})} \boldsymbol{\Phi}^{\mathrm{T}})^{\dagger} (\mathbf{s} - \boldsymbol{\Phi} \boldsymbol{\mu}_{\mathbf{x}}^{(\mathbf{k})}) \right), \qquad (9)$$

where

$$\tilde{\pi}_k = \frac{\pi_k \mathcal{N}(\mathbf{s}; \boldsymbol{\Phi}\boldsymbol{\mu}_{\mathbf{x}}^{(\mathbf{k})}, \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\mathbf{x}}^{(\mathbf{k})}\boldsymbol{\Phi}^{\mathrm{T}})}{\sum_{\ell=1}^{K} \pi_\ell \mathcal{N}(\mathbf{s}; \boldsymbol{\Phi}\boldsymbol{\mu}_{\mathbf{x}}^{(\ell)}, \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\mathbf{x}}^{(\ell)}\boldsymbol{\Phi}^{\mathrm{T}})}. \qquad (10)$$

Once obtained the conditional mean estimate $\hat{\mathbf{x}} = [\hat{\mathbf{a}}^{\mathrm{T}}, \hat{\mathbf{p}}^{\mathrm{T}}]^{\mathrm{T}}$, the vectors $\hat{\mathbf{a}}$ and $\hat{\mathbf{p}}$ can be used to build estimates of the A2 and P2 components $\hat{a}_i(n)$ and $\hat{p}_i(n)$. In particular, overlapping vectors can be extracted from the recorded heart sound signals, thus implying that multiple vectors $\hat{\mathbf{a}}$ and $\hat{\mathbf{p}}$ may contain estimates for a given sample of $a_i(n)$ or $p_i(n)$. In this case, sample estimates are obtained by simple averaging.

3. Numerical results

The performance of the proposed method in separating A2 and P2 components is compared with that provided by the method in [6]. Performance is evaluated in terms of the normalized root mean-squared error (NRMSE) achieved in the reconstruction of A2 and P2 components.

3.1. Signal model

The tests reported in this work consider synthetic heart sounds generated according to the waveform model described in [5]. In particular, the A2 and P2 components of the i-th S2 sound of a given recording are modeled as windowed sinusoids, with Gaussian windows, i.e.,

$$a_i(n) = w_i^{(a)}(n) \cos(2\pi f^{(a)} \frac{n}{F_s} + \theta^{(a)} - \mu_i^{(a)}) + z_i^{(a)}(n)$$
$$p_i(n) = w_i^{(p)}(n) \cos(2\pi f^{(p)} \frac{n}{F_s} + \theta^{(p)} - \mu_i^{(p)}) + z_i^{(p)}(n),$$

where

$$w_i^{(a)}(n) = \alpha^{(a)} \exp\left(\frac{-(n/F_s - \mu_i^{(a)})^2}{2\beta^{(a)}}\right)$$
$$w_i^{(p)}(n) = \alpha^{(p)} \exp\left(\frac{-(n/F_s - \mu_i^{(p)})^2}{2\beta^{(p)}}\right),$$

and $z_i^{(a)}(n)$ and $z_i^{(p)}(n)$ represent additive white Gaussian noise (AWGN) components which account for deviations of the recorded signals from the theoretical model adopted. As described in [5], the amplitude parameters $\alpha^{(a)}$ and $\alpha^{(p)}$ are drawn from normal and log-normal distributions, respectively. The spreading parameters $\beta^{(a)}$ and $\beta^{(p)}$ are drawn from log-normal and uniform distributions, respectively. The frequency parameters $f^{(a)}$ and $f^{(p)}$ are drawn from log-normal distributions, whereas the phase parameters $\theta^{(a)}$ and $\theta^{(p)}$ are drawn from uniform distributions. The delay parameters $\mu_i^{(a)}$ are drawn from a log-normal distribution. The parameters defining such distributions are inferred by fitting such distributions to the data obtained from a cohort including 150 subjects [5]. Finally, the delay parameters $\mu_i^{(p)}$ are drawn from uniform distributions, whose parameters were chosen according to the statistics of the A2-P2 split $\Delta_i = \mu_i^{(p)} - \mu_i^{(a)}$ considered in each test.
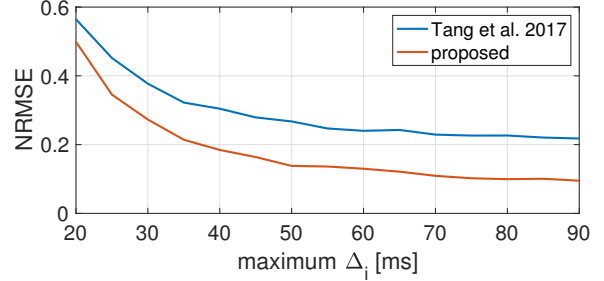

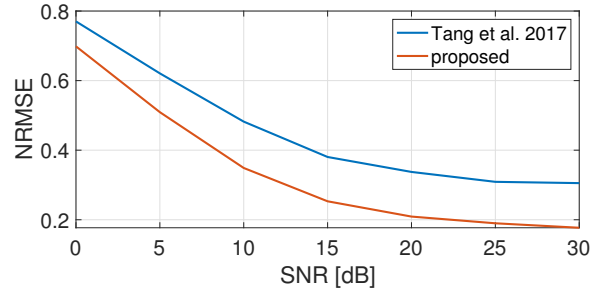Figure 1. NRMSE versus maximum Δ_i. SNR $= 30$ dB.

Figure 2. NRMSE versus SNR $= 30$ dB. Minimum $\Delta_i = 10$ ms, maximum $\Delta_i = 40$ ms.

3.2. Results

Synthetic heart sound signals are generated with sampling frequency $F_s = 1$ kHz. Vectors of interest are extracted from such signals according to the procedure described in Section 2.1 by picking $N_t = 8$ samples from $N_b = 8$ consecutive S2 sounds. Such vectors are overlapping in the time domain, as they are extracted from heart sound signals with a stride of 2 samples.

A training dataset is formed by generating 100 heart sound recordings, each containing 16 heartbeats. From this data, a joint GMM with $K = 20$ Gaussian components is inferred, which models the vectors $\mathbf{x} = [\mathbf{a}^{\mathrm{T}}, \mathbf{p}^{\mathrm{T}}]^{\mathrm{T}}$ containing samples from the A2 and P2 components. Then, the proposed separation algorithm and the algorithm described in [6] are applied to further 400 heart sounds, each containing 16 heartbeats, which are generated independently from the training data.

First, the performance of the considered separation algorithms is tested against different values of A2-P2 splits. The NRMSE in recovering A2 and P2 components from S2 sounds is reported in Fig. 1, when the split is distributed uniformly between 10 ms and values ranging from 20 ms to 90 ms. As expected, source separation is less reliable when the split is lower, due to the increased time-domain overlap between the A2 and P2 components. On the other hand, the proposed approach is shown to outperform significantly the method described in [6] for all splits considered.

Then, in order to evaluate the robustness of the proposed method against deviations of the recorded heart sound signals from the adopted joint GMM, simulations have been run by training the GMM with sounds with a given amount of distortion (corresponding to an SNR value of 30 dB) and testing the model with heart sounds with distortions with different amplitude, corresponding to SNR values ranging from 0 to 30 dB. The corresponding NRMSE values are reported in Fig. 2, which shows how the proposed method guarantees more accurate reconstructions with respect to the algorithm in [6] for all the different distortion values considered.

In general, using joint GMM priors to model portions of fundamental heart sound components extracted from successive heartbeats is shown to have a considerable potential in solving the inverse problem associated to the separation of A2 and P2 components from the observation of S2 recordings. This is mainly motivated by the ability of GMM priors in modeling different complex signals and in discriminating efficiently signals that occupy low-dimensional manifolds formed by the union of subspaces with little intersection [10].

On the other hand, the main limitation of the proposed approach is represented by the need of a training dataset containing separated A2 and P2 components to infer the parameters of the joint GMM prior. This requirement turns out particularly challenging when considering the application of the proposed approach to real-world heart sound recordings, due to the scarcity of available datasets with annotations discriminating the different components of S2 sounds.

4.   Conclusion

A novel method to separate A2 and P2 components from S2 recording has been presented. The method, which is data-driven, models segments of A2 and P2 components extracted from successive S2 sounds via a joint GMM. The proposed method is shown to recover A2 and P2 components more precisely than other algorithms previously presented in the literature. Also the proposed approach shows promising results regarding its robustness against small A2-P2 splits and against signal deviations from the trained model.

Future work encompasses the test of the proposed algorithm with real heart sound data recordings.

Acknowledgements

References

[1] Xu J, Durand LG, Pibarot P. A new, simple, and accurate method for non-invasive estimation of pulmonary arterial pressure. Heart 2002;88(1):76–80.

[2] Smith R, Ventura D. A general model for continuous non-invasive pulmonary artery pressure estimation. Computers in biology and medicine 2013;43(7):904–913.

[3] Comon P, Jutten C. Handbook of Blind Source Separation: Independent Component Analysis and Applications. 1st edition. Orlando, FL, USA: Academic Press, Inc., 2010. ISBN 0123747260, 9780123747266.

[4] Bobin J, Starck JL, Fadili J, Moudden Y. Sparsity and morphological diversity in blind source separation. IEEE Transactions on Image Processing Nov 2007;16(11):2662–2674.

[5] Sæderup RG, Hoang P, Winther S, Bøttcher M, Struijk J, Schmidt S, Østergaard J. Estimation of the second heart sound split using windowed sinusoidal models. Biomedical Signal Processing and Control 2018;44:229–236.

[6] Tang H, Chen H, Li T. Discrimination of aortic and pulmonary components from the second heart sound using respiratory modulation and measurement of respiratory split. Applied Sciences 2017;7(7):690.

[7] Renna F, Calderbank R, Carin L, Rodrigues MRD. Reconstruction of signals drawn from a gaussian mixture via noisy compressive measurements. IEEE Transactions on Signal Processing May 2014;62(9):2265–2277.

[8] Renna F, Oliveira J, Coimbra MT. A data-driven feature extraction method for enhanced phonocardiogram segmentation. In 2017 Computing in Cardiology (CinC). Sep. 2017; 1–4.

[9] Renna F, Wang L, Yuan X, Yang J, Reeves G, Calderbank R, Carin L, Rodrigues MRD. Classification and reconstruction of high-dimensional signals from low-dimensional features in the presence of side information. IEEE Transactions on Information Theory Nov 2016;62(11):6459–6492.

[10] Sabetsarvestani Z, Renna F, Kiraly F, Rodrigues MRD. Source separation in the presence of side information: Necessary and sufficient conditions for reliable de-mixing. In 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP). Nov 2018; 351–355.

[11] Bishop C. Pattern Recognition and Machine Learning (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

Address for correspondence:

Name: Francesco Renna
Full postal address: Rua do Campo Alegre 1021/1055,
4169-007 Porto, Portugal
E-mail address: frarenna@dcc.fc.up.pt