Multimodal Sensor Fusion of Cardiac Signals via Blind Deconvolution: A Source-Filter Approach

Christoph Hoog Antink, Christoph Brüser, Steffen Leonhardt

Philips Chair for Medical Information Technology, RWTH Aachen University, Germany

Abstract

Sensor fusion is a growing field within the medical signal processing community. Traditionally, it is performed implicitly by the physician when diagnosing the state of a patient from various measurement modalities such as electrocardiography (ECG), arterial blood pressure (ABP) or photoplethysmography (PPG). These may represent different physical quantities like voltage, pressure or scattering properties and are modulated by various physiological conditions and artifacts. Still, they originate from a single source, the heart. In signal processing, this is known as a single-input-multiple-output (SIMO) system and several approaches to estimate the source are known. In this paper, a blind deconvolution approach is chosen and adapted for physiological signals. The feasibility and robustness is shown using simulated data. Moreover, the approach is validated on real data recorded in a polysomnography setting, fusing PPG and Ballistocardiography (BCG).

1. Introduction

The purpose of cardiac monitoring in general is, technically speaking, to infer the "state" of the heart or, more general, the state of the patient. Typical parameters are the heart rate (HR), the heart rate variability (HRV) or stroke volume (SV), to name only a few. Various modalities exist, ranging from classical clinical methods like electrocardiography (ECG), invasive blood pressure (IBP) measurements or photoplethysmography (PPG) to unobtrusive monitoring techniques like ballistocardiography (BCG), capacitive electrocardiography (cECG) or photoplethysmography imaging (PPGI). Especially those of the latter category are very sensitive to noise and motion artifacts, but even the classical modalities can be severely contaminated with artifacts. To overcome this, single- [1] and multi-modal [2,3] sensor-fusion is seeing increasing attention, another example being the current "PhysioNet/CinC Challenge 2014".

In this article, we propose to interpret the generation of physiological signals related to the cardiac activity as a source-filter problem. In speech processing, such a model is very common [4], where the vocal chords can be considered as a time-varying quasi-periodic signal generator and the vocal tract as a time-varying filter. The signal generator in this single-input single-output model creates an actual physical signal (air pressure variations) that is modulated with a physical filter (acoustic impedance of the vocal tract). In contrast, the source signal considered in our model can be purely virtual, see Figure 1: The single

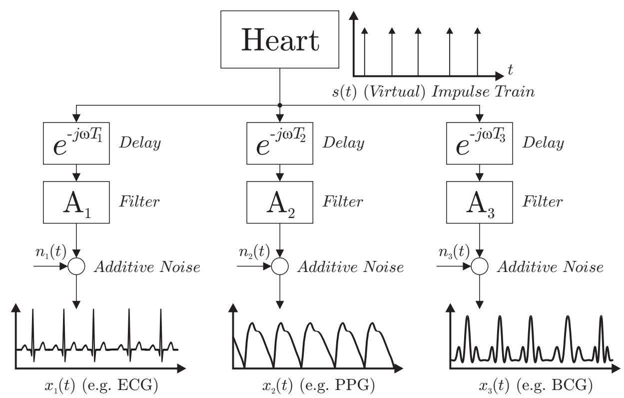

Figure 1. Source-Filter Model for cardiac signals: The heart creates a (virtual) impulse train that is delayed, filtered and contaminated with noise to create observed sensor outputs.

source, i.e. the heart, generates a quasi-periodic signal, in this example a train of impulses. In the case of m sensor channels, this signal is filtered with m delay- and transfer functions that generate the actual physical representations or (including the sensor transfer functions and adding noise) the sensor outputs. In the general case, these transfer functions will be non-linear and time-varying. However, in this proof-of-concept, we will consider the linear, time-invariant case. This will allow us to apply techniques from the domain of blind deconvolution used for single-input-multiple-output (SIMO) systems to this problem. In [5], an algorithm was presented that allows the estimation of a source signal and m transfer functions from m observations without imposing limitations on the source signal. In this paper, this algorithm is adapted and its feasibility is analyzed. The paper is structured as follows: In Section 2, the original algorithm and our modifications are presented. In 3, it is demonstrated on synthetic data. The algorithm's performance is tested on an example of a polysomnographic dataset in Section 4. The paper is concluded with results and conclusions drawn in Section 5.

2. SIMO Deconvolution Algorithm

The general idea of the algorithm presented in [5] is to perform the deconvolution process iteratively in the Fourier domain, where the time-domain convolution is represented by a multiplication: Let $\vec{x}(t) \in \mathbb{R}^{m\times 1}$ be the observed, discrete-time signal, where m is the number of channels and $t \in \{1, ..., T\}$. If $s(t)$ is the source signal and $\vec{n}(t) \in \mathbb{R}^{m\times 1}$ signifies additive noise,

$$\vec{x}(t) = \sum_{\tau=0}^{q} \vec{a}(\tau)s(t-\tau) + \vec{n}(t). \quad (1)$$

Here, $\vec{a}(\tau) \in \mathbb{R}^{m\times 1}$ are the filter coefficients, i.e., the impulse response of each of the m channels, q is the order of the filters. Transforming into the Fourier domain results in

$$\vec{X}(\omega) = \sum_{\tau=0}^{q} \vec{a}(\tau)S(\omega)e^{-j2\pi\omega\tau/T} + \vec{N}(\omega) \quad (2)$$

where $\vec{X}(\omega)$, $S(\omega)$ and $\vec{N}(\omega)$ are the Fourier transforms of their time-domain counterparts. Note that the filter coefficients are not transformed but the delay is expressed explicitly by the term $e^{-j2\pi\omega\tau/T}$. This can be rewritten in matrix notation as

$$\vec{X}(\omega) = \mathbf{A} \cdot \vec{E}(\omega) \cdot S(\omega) + \vec{N}(\omega), \quad (3)$$

with the Fourier-domain delay-vector

$$\vec{E}(\omega) = \left(1, e^{-j2\pi\omega/T}, ..., e^{-j2\pi\omega q/T}\right)^T \quad (4)$$

and the matrix of filter coefficients

$$\mathbf{A} = \left(\vec{a}(0), \vec{a}(1), ..., \vec{a}(q)\right). \quad (5)$$

Let $(\cdot)^T$ be the transpose, $(\cdot)^H$ the Hermitian transpose and $(\cdot)^*$ the conjugation operator. If the filter coefficients were known and the noise is zero-mean Gaussian, a source signal that minimizes a quadratic cost function can be obtained through

$$S_{est}(\omega) = \left(\vec{E}^H(\omega)\mathbf{A}^T\mathbf{A}\vec{E}(\omega)\right)^{-1} \vec{E}^H(\omega)\mathbf{A}^T\vec{X}(\omega). \quad (6)$$

If the source signal was known, the optimal filter coefficients could be determined accordingly with

$$\mathbf{A}_{est} = \left(\mathbf{X}^* \cdot \mathbf{ES}^T + \mathbf{X} \cdot \mathbf{ES}^H\right) \cdot \left(\mathbf{ES}^* \cdot \mathbf{ES}^T + \mathbf{ES} \cdot \mathbf{ES}^H\right)^{-1} \quad (7)$$

with the matrices

$$\mathbf{ES} = \left(\vec{E}(1) \cdot S(1), \vec{E}(2) \cdot S(2), ..., \vec{E}(T) \cdot S(T)\right) \quad (8)$$

and

$$\mathbf{X} = \left(\vec{X}(1), \vec{X}(2), ..., \vec{X}(T)\right). \quad (9)$$

For the derivation of Equations 6 and 7, see [5]. In the original algorithm, the matrix of filter coefficients $\mathbf{A}_{est}$ is randomly initialized, and an optimal source signal S_{est} is determined. Iteratively, filter coefficients and source signal are optimized, a process called *Successive Substitution Method*. This approach was demonstrated to converge, granted that $T \gg q$ and $m \geq 2$. However, the original demonstration showed some properties that do not transfer to our case: First, the source signal S used for demonstration was a "woman's voice [...] sampled at 65536 points with the frequency 22050 Hz", which, judging from the figures provided, was non-stationary. Here, we assume the source to be quasi-periodic, which results in a much more sparse spectrum. Next, the observed signal $\vec{x}(t)$ with $T = 65536$ were created from the source with random filters of order $q = 2$, representing an impulse response of only $1.3605 \cdot 10^{-04}$ seconds duration. While this is probably reasonable for some technical systems, considering, for example, the ECG to be an "impulse response" to a train of impulses, a much higher filter order q has to be assumed. To keep the ratio of T and q, a much longer signal had to be considered in consequence. This would increase computational cost dramatically and stands in contrast to the assumption of local time invariance.

To improve convergence in our less well-posed scenario, the random initialization step was performed multiple (here 100) times and two iterations of the algorithm were conducted. The best result was then chosen as the starting point for the final iterations. This improved convergence speed significantly. To increase efficiency, the symmetry of the spectrum was exploited, cutting computational cost approximately in half.

A significant factor when examining real cardiac signals obtained by different sensors is delay: If, for example, ECG and ABP are considered, there is a delay of roughly 200 ms between the QRS-peak and the peak in the pulse wave. If this is to be represented by an FIR-filter, the order of filter has to be increased accordingly, worsening the condition of the problem. To overcome this, a pre-alignment step is introduced. One of the input signals is arbitrarily chosen as reference and the cross-correlation between it and the remaining signals is computed. The lag-index of the maximum absolute cross-correlation is used for alignment with the reference signal.

3. Simulative Demonstration

To evaluate the algorithm, a synthetic test was conducted. The source signal $s(t)$ was created as a train of 80 impulses of unit height with varying distances. The distance between pulses was normally distributed with a mean of 25 and a variance of 5 samples, simulating a mean heart-rate of 60 beats per minute sampled at 25 Hz. The number of channels was set to $m = 5$ and the order of the filter was set to $q = 13$. Next, the matrix $\mathbf{A}$ was randomly initialized and normalized such that

$$\sum_{j=0}^{q} \mathbf{A}(i,j) = 1 \qquad \text{for } i \in (1 \ldots m). \tag{10}$$

Via convolution, the observed signals $\vec{x}(t)$ are calculated. Next, normally distributed noise with variance $\sigma = 0.5$ was added to each channel. To mimic an outlier affecting all channels simultaneously, a single sample at a random position was altered,

$$\vec{x}(654) = (10, 10, 10, 10, 10)^T. \tag{11}$$

Figure 2 shows the first 800 samples of $s(t)$ and the first two observed channels $x_1(t)$ and $x_2(t)$. Without pre-alignment, deconvolution is performed. Figure 3 shows the resulting estimated source signal $s_{est}(t)$ and the original source. It can be seen that the impulses of the original source are recovered well via deconvolution. One can also see that s_{est} shows more noise in the vicinity of the outlier. However, the position of the actual impulse at $t = 650$ is not altered and no additional impulse with high amplitude is introduced.

4. Application to Polysomnographic Data

Next, the concept was tested on real multimodal patient data. For this, a 160 second excerpt from a polysomnography recording was used. For computational reasons the data was re-sampled to 25 Hz. The channels PPG and BCG were selected as observed signals $x_1(t)$ and $x_2(t)$. The signals were pre-filtered to be mean-free and of unit variance, the aforementioned pre-alignment was performed. Experimentally, the filter-order was determined to $q = 20$, corresponding to an impulse-response duration of 0.84 seconds. In Figure 4, the first 20 seconds of the observed signals and the estimated source signal are shown. It can be noticed that not an impulse-train but a smooth, almost sinusoidal signal was estimated, seemingly containing a respiratory

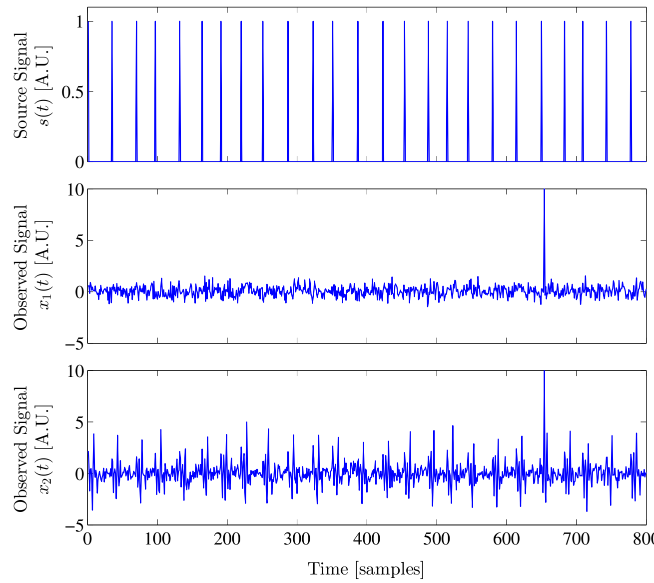

Figure 2. Synthetic Example: Original source signal and first two observed channels containing Gaussian noise and a random outlier at $t = 654$.

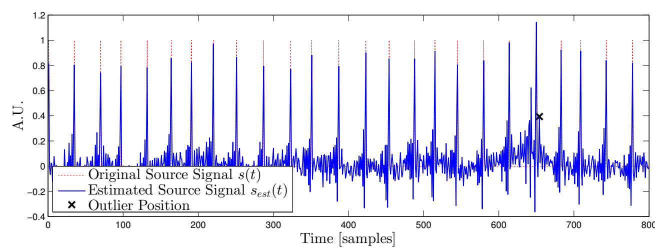

Figure 3. Reconstructed (synthetic) source signal.

and a cardiac component. As a most simple approach, the signal was split into a high-pass (HP, $f > 1$ Hz) and a low-pass (LP, $f < 0.5$ Hz) component. In Figure 5, these components are shown together with their respective reference signals ECG and nasal flow. Via peak-detection, the beat-to-beat and breath-to-breath intervals were analyzed. For the cardiac components, the results are presented in Figure 6: A close match of both curves (RMSE = 42.9 ms) can be examined, showing beat-to-beat accuracy. Very promising results were obtained for the breath-to-breath interval analysis as well (RMSE = 265.7 ms).

5. Results and Conclusion

First of all, the general feasibility of the approach could be demonstrated. Even though the spectrum of the source signal was relatively sparse and the ratio of filter coefficients and signal length was worse, a general deconvolution strategy [5] could be adapted to a medical instrumen-

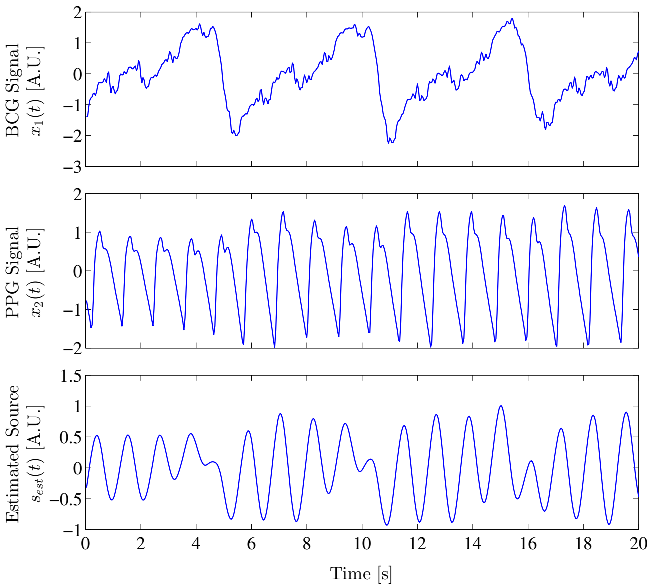

Figure 4. Real Data: BCG channel, PPG channel and estimated source signal.

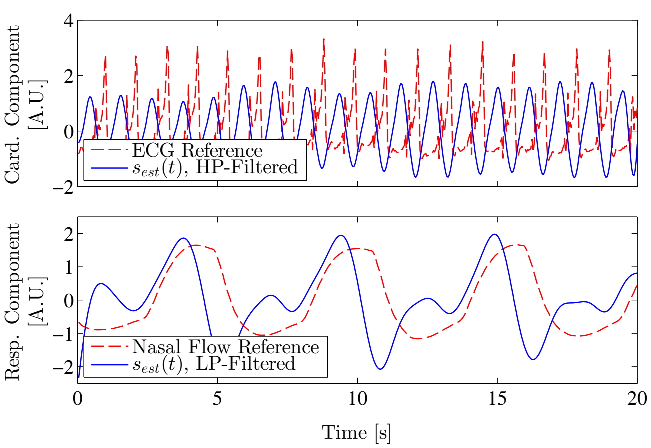

Figure 5. Real Data: Cardiac and respiratory component of the estimated source signal with respective reference signals, ECG and nasal flow.

tation problem, thereby supporting the presented source-filter model for physiological signals. For simulated data, the approach proved to be robust to Gaussian noise and even outliers of high amplitude affecting all channels. This is reasonable since the deconvolution is performed in the spectral domain, where a single time-domain impulse is represented by a flat spectrum. At the same time this is a realistic scenario in unobtrusive sensing, where motion artifacts can affect multiple channels. Finally, the feasibility could be demonstrated on real polysomnography data as well, even in the presence of two sources (heart and lung)

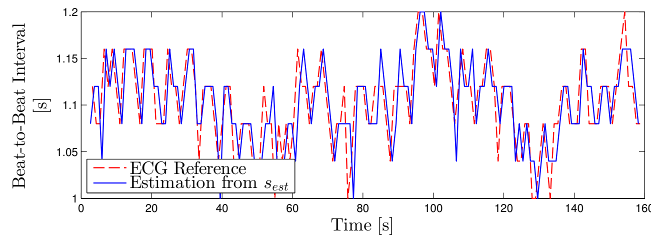

Figure 6. Real Data: Beat-to-Beat analysis of ECG signal and estimated source signal, RMSE = 42.9 ms

in the observed signals.

As a next step, the approach has to be tested extensively on various measurement modalities and medical conditions. Also, the separation of individual sources from the estimated source signal has to be improved. Hereby, one has to keep in mind that scaling and even the sign of source signal and filters are not unique. Still, the analysis of the filter coefficients over time could provide diagnostic information. Moreover, using the estimated source signal and filter coefficients, one can predict individual observed channels. This could be used for data compression or, in the event of a large residual as an "event-indicator", highlighting interesting segments in the data that could hint at medical conditions or artifacts.

References

[1] Brueser C, Kortelainen J, Winter S, Tenhunen M, Parkka J, Leonhardt S. Improvement of Force-sensor-based Heart Rate Estimation Using Multi-channel Data Fusion. IEEE Journal of Biomedical and Health Informatics 2014;Preprint(99).

[2] Silva I, Lee J, Mark RG. Signal quality estimation with multichannel adaptive filtering in intensive care settings. IEEE Transactions on Biomedical Engineering September 2012; 59(9):2476–2485.

[3] Wartzek T, Brueser C, Walter M, Leonhardt S. Robust Sensor Fusion of Unobtrusively Measured Heart Rate. IEEE Journal of Biomedical and Health Informatics 2014;18(2):654–660.

[4] Huang X, Acero A, Hon HW. Spoken language processing, volume 18. Englewood Cliffs: Prentice Hall, 2001.

[5] He Z, Xie S, Fu Y. A new blind deconvolution algorithm for SIMO channel based on neural network. In Proceedings of the Fourth International Conference on Machine Learning and Cybernetics. 2005; 3602–3606.

Address for correspondence:

Christoph Hoog Antink
Chair for Medical Information Technology
Helmholtz-Institut, RWTH Aachen
Pauwelsstr. 20 / D-52074 Aachen / Germany
hoog.antink@hia.rwth-aachen.de