Peer-reviewed version available at *Technologies* **2018**, *6*, 26; doi:doi:10.3390/technologies6010026

Article

A Low-Complexity Model-Free Approach for Real-Time Cardiac Anomaly Detection Based on Singular Spectrum Analysis and Nonparametric Control Charts

Michael Lang [1,*]

[1] Graduate School of Excellence Computational Engineering, Technische Universtät Darmstadt, Dolivostraße 15, 64293 Darmstadt, Germany; Tel.: +49 6151 1624401, Fax: 49 6151 1624404

[*] Correspondence: michael.lang@ieee.org

Abstract: While the importance of continuous monitoring of electrocardiographic (ECG) or photoplethysmographic (PPG) signals to detect cardiac anomalies is generally accepted in preventative medicine, there remain major barriers to its actual widespread adoption. Most notably, current approaches tend to lack real-time capability, exhibit high computational cost, and be based on restrictive modeling assumptions or require large amounts of training data. We propose a lightweight and model-free approach for the online detection of cardiac anomalies such as ectopic beats in ECG or PPG signals based on the change detection capabilities of Singular Spectrum Analysis (SSA) and nonparametric rank-based cumulative sum (CUSUM) control charts. The procedure is able to quickly detect anomalies without requiring the identification of fiducial points such as R-peaks and is computationally significantly less demanding than previously proposed SSA-based approaches. Therefore, the proposed procedure is equally well suited for standalone use and as an add-on to complement existing (e.g. heart rate (HR) estimation) procedures.

Keywords: nonparametric change point detection; singular spectrum analysis; cumulative sums; ecg; ppg; arrhythmias; cardiac monitoring

1. Introduction

The ubiquity of powerful smartphones and other smart devices, which nowadays incorporate a plethora of advanced sensing capabilities, has led to an increasing trend in the consumer sphere to continuously gather and evaluate physiological signals [1,2]. In particular, cardiovascular parameters such as heart rate and pulse rate (PR), extracted respectively from measurements of myocardial electrical potentials through ECG [3–6] and from measurements of volumetric changes of blood perfusion during cardiac cycles by optical means through PPG [7–9] are being recorded and analyzed in apps, fitness trackers and the like with great potential benefits for public health [1,10–12].

Virtually all of said consumer-oriented apps and devices fall into the category of fitness and well-being products, thereby avoiding the substantial burden of having to comply with requirements imposed by medical devices regulatory frameworks [13,14], which is reflected in often notoriously inaccurate and unreliable results [15,16]. Moreover, functionality is usually limited to providing estimates of the average heart rate.

While low-resolution averaged HR estimates may be of some use in fitness and well-being scenarios, from a clinical perspective the detection of sudden changes in the signal structure are of utmost importance. ECG recordings from a healthy heart are characterized by a *sinus rhythm*, wherein the normal cardiac cycle begins with an action potential in the *sinoatrial (SA) node*, located in the *right atrium*, which propagates and depolarizes neighboring cells. The depolarization of the *SA node* spreads rapidly throughout both atria, specifically to the *left atrium* through *Bachmann's bundle* and through internodal pathways in the *right atrium* to the *atrioventricular (AV) node*. Full depolarization gives rise to the P wave, which initiates atrial contraction. From the AV node, excitation is further propagated after an initial delay of about 100 ms through the *bundle of His*, which splits up into the *right bundle branch* and the *left bundle branch*, initiating respectively the depolarization of the *right and left ventricle*, yielding to the *QRS-complex* which ends with completely depolarized and contracting ventricles.

Ventricular repolarization following the contraction eventually results in the T-wave and concludes the normal cardiac cycle. Note that both left and right bundle branches eventually differentiate into a large number of *Purkinje fibers*, the repolarization of which is thought to occasionally result in an additional U-wave [3].

Deviations from the normal sinus rhythm are referred to as *arrhythmias* and comprise a large number of specific arrhythmias (*see, e.g.* [17]). Heart rhythms exhibiting variations in timing such as those that are either below 60 beats per minute (bpm) or above 100 bpm as well as rhythms disrupted by changes in the morphology, e.g. due to ectopic beats (i.e. heart beats whose origin is different from and outside of the region typically responsible for impulse generation, namely the SA node) such as *premature ventricular contractions* (PVCs) or *premature atrial contractions* (PACs) all qualify as arrhythmias. Furthermore, they all tend to induce distinctive changes in the ECG signal thereby allowing for a change detection approach which must not necessarily be based on templates corresponding to the various arrhythmia-induced changes [18]. Also, many arrhythmias, though usually not acutely life-threatening, are paroxysmal and asymptomatic and therefore likely to go unnoticed for long periods of time, which carries the risk of exacerbation and possibly the development of more serious types of arrhythmias [3,17,18]. Continuous outpatient cardiac monitoring through wearable devices is commonly accepted as a promising approach to tackle this issue and improve treatment outcome while at the same time lowering overall healthcare costs [1,10–12,18,19].

While the automatic monitoring of ECG signals has been researched for a couple of decades and various algorithms have been proposed and implemented, the shift towards continuous outpatient monitoring through low power wearable devices introduces numerous additional and challenging requirements such as real-time capability, the ability to cope with rather noisy and low-quality signals with various artifacts and harsh constraints on computational complexity and power consumption [18–20]. Various approaches for the automatic online detection of cardiac arrhythmias have been proposed in the literature. Machine Learning approaches have been adopted by numerous authors [21,22], as have Wavelet [22,23], Artificial Neural Network (ANN) [24–26], and decision tree [27] based approaches. For a more detailed review, the reader is referred to some recent review papers [18,28]. Typically most of these approaches require the extraction of certain features from the signal, e.g. the location of the QRS-complexes [20,27,29,30] or the R-peaks [22,23,31–33], commonly performed using the algorithm proposed by Pan and Tompkins [34] or variations thereof [18,28], and are therefore inherently vulnerable to inaccuracies in the initial estimation of these fiducial points. Moreover they tend to be based on rather restrictive modeling assumptions and/or require large amounts of training data, which makes them hard to reconcile with the requirements arising in real world outpatient monitoring scenarios.

We propose a lightweight and model-free approach for the online detection of cardiac anomalies such as ectopic beats in ECG or PPG signals based on the change detection capabilities of Singular Spectrum Analysis (SSA) and nonparametric rank-based cumulative sum (CUSUM) control charts. The procedure is able to quickly detect anomalies without requiring templates, extensive training data sets or the identification of fiducial points such as R-peaks and is computationally significantly less demanding than previously proposed SSA-based approaches.

The proposed method is essentially composed of two consecutive steps: an SSA-based algorithm is sequentially applied to the observed data to construct viable test statistics that reflect potential changes in the cardiac signal, and these statistics are then monitored using distribution-free CUSUM-type control charts.

The use of SSA in a sequential framework as a means for change detection as introduced in Section 2 is based on works by Moskvina and Zhigljavsky, discussed in detail in [35] and in a more condensed fashion in [36], although it should be noted that the concept was earlier already described in [37]. Said algorithm has successfully been applied to various real world detection problems, e.g. anomaly detection in Cognitive Radio Networks [38], smart power grids [39], software engineering [40] and change point detection for complex-valued time series [41]. In the biomedical context it has shown to be useful for the identification of freezing of gait in patients with Parkinson's disease [42], the detection of anomalies in periodic biosignals such as ECGs [43] as well as other biomedical applications [44].

Building on the prior art outlined in Section 2, we introduce a novel SSA-based change detection

procedure (l-SSA-CPD) that exhibits very reasonable performance characteristics while at the same time drastically reducing the computational burden. As shall be shown in Section 3, this is accomplished by modifying the conventional SSA-based change detection algorithm such that the computationally expensive task of computing the Singular Value Decomposition (SVD) is only performed at the very beginning instead of each time a new data point becomes available. Furthermore, the proposed procedure uses more elaborate test statistics that take into account the information derived from the angle between data vectors representing new observations and the subspace representing the signals characteristics as well as the euclidean distances. Lastly, our procedure differs from previous approaches also in that rank-based control limits and the reinitialization of control charts after an anomaly has been detected are used. A performance evaluation of l-SSA-CPD using ECG and PPG records from the publicly available Physionet Challenge 2015 training database (PC15) [45,46] will be presented in Section 4 and followed by a short discussion in Section 5 which concludes this paper.

2. Singular Spectrum Analysis

2.1. Fundamentals of Singular Spectrum Analysis

Singular Spectrum Analysis is a technique of time series analysis and can be interpreted as belonging to the general class of Principal Component Analysis (PCA) methods. SSA has become a standard tool in meteorology and climatology but is mostly unknown outside of those disciplines. Golyandina et al. [37] attribute this to the nature of SSA being more a technique of multivariate geometry than of statistics. According to their representation, SSA should rather be seen as an exploratory, model-building tool than a confirmatory procedure. In essence, SSA can be seen as the application of PCA to the so-called trajectory matrix (obtained directly from the original time series) with the subsequent attempt to reconstruct the original series. Prior to proceeding to SSA for change detection a short introduction to the basic SSA algorithm appears in order.

2.1.1. Basic SSA Algorithm

Consider N observations $\mathbb{X}_N = (x_1, \ldots, x_N)$ of a univariate time series and an integer M $(1 < M << N)$ commonly referred to as *window length*, *lag-integer* or *embedding dimension*. The basic SSA algorithm is commonly described as being composed of the following four stages (*see, e.g.* [37,44,47]):

1. Embedding: $\mathbb{X}_N = (x_1, \ldots, x_N) \rightarrow \mathbf{X} \in \mathbb{R}^{M \times K}$

   A so-called *trajectory matrix* $\mathbf{X}$ is constructed by mapping $\mathbb{X}_N$ into a sequence of $K = N - M + 1$ lagged column vectors $X_j = (x_j, \ldots, x_{j+M-1})^T$, $j = 1, \ldots, K$ of size M, yielding

$$\mathbb{X}_N = (x_1, x_2, \ldots, x_N) \rightarrow \mathbf{X} = \begin{bmatrix} x_{n+1} & x_{n+2} & \cdots & x_{n+K} \\ x_{n+2} & x_{n+3} & \cdots & x_{n+K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n+M} & x_{n+M+1} & \cdots & x_{n+N} \end{bmatrix} \tag{1}$$

   Notice the *Hankel-structure* of $\mathbf{X} = (x_{ij})_{i,j=1}^{M,K}$, i.e. $\mathbf{X}$ has equal elements on the anti-diagonals $i + j = \text{const}$.

   One can think of $\mathbf{X}$ as multivariate data with M characteristics and K observations and accordingly X_j of $\mathbf{X}$ as vectors in the M-dimensional space $\mathbb{R}^M$.

2. Singular Value Decomposition of $\mathbf{X}$

   Taking the SVD of $\mathbf{X}$ decomposes the trajectory matrix into its orthogonal bases and yields a collection of M eigenvalues and eigenvectors. Let $\lambda_1 \geq \cdots \geq \lambda_M \geq 0$ and $U_1, \ldots, U_M$ denote, respectively, the eigenvalues and eigenvectors of $\mathbf{X}\mathbf{X}^\mathbf{T}$ and the *rank* of $\mathbf{X}$ be denoted as $d = \max(i, \text{ such that } \lambda_i > 0)$. The SVD of $\mathbf{X}$ can then be rewritten as the sum of d *elementary matrices*

$$\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_d \tag{2}$$

with matrices $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ being of rank 1 and $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$.

Note that V_i are the eigenvectors of $\mathbf{X}^T\mathbf{X}$ and $\left(\sqrt{\lambda_i}, U_i, V_i\right)$ the *eigentriples* of the SVD in Eq. (2).

Also note that due to the symmetry of left and right singular vectors, the SVD of trajectory matrices obtained with window length M and $K = N - M + 1$ are equivalent. Accordingly, one can impose the limitation $M \leq N/2$ on the window length since there is no additional benefit in using a larger window (*see, e.g.* [47] at 47, [37] at 69).

3. Eigentriple Grouping

In order to separate the signals of interest from noise and artifacts, the third stage of basic SSA aims to find particular disjoint subsets of the set of indices $\{1, \ldots, d\}$ such that the respective systems of eigenvectors span the subspaces associated with the different signal components.

Consider the task of separating a signal of interest from unwanted noise. One then looks for a certain subset of indices $I = \{i_1, \ldots, i_l\}, l < d \leq M$ that span an l-dimensional subspace in $\mathbb{R}^M$, denoted as $\mathcal{L}_I \subset \mathbb{R}^M = span\{U_I\} = span\{U_{i_1}, \ldots, U_{i_l}\}$. Analogously, the remaining eigentriples with $\bar{I} = \{i_1, \ldots, i_d\} \setminus I$ span the noise subspace $\mathcal{L}_{\bar{I}} \subset \mathbb{R}^M = span\{U_{\bar{I}}\}$.

The trajectory matrix component $\mathbf{X}_I$ corresponding to the subset I of eigentriples associated with the signal of interest is then

$$\mathbf{X}_I = \mathbf{X}_{i_1} + \cdots + \mathbf{X}_{i_l} \tag{3}$$

and the component $\mathbf{X}_{\bar{I}}$ corresponding to the subset $\bar{I} = \{i_1, \ldots, i_d\} \setminus I$ associated with the remainder of the observed signal is

$$\mathbf{X}_{\bar{I}} = \sum_{i \in \bar{I}} \mathbf{X}_i \tag{4}$$

such that

$$\mathbf{X} = \mathbf{X_I} + \mathbf{X}_{\bar{I}} = \sum_{i \in I} \mathbf{X}_i + \sum_{i \in \bar{I}} \mathbf{X}_i \tag{5}$$

In the case of separability (*see, e.g.* [47] at 17), the contribution of $\mathbf{X}_I$ to the entire observed signal $\mathbf{X}$ is represented by the respective share of eigenvalues $\sum_{i \in I} \lambda_i / \sum_{i=1}^{d} \lambda_i$.

4. Diagonal Averaging

For perfectly separable components, all matrices in the expansion of Eq. (5) are Hankel matrices. For real world problems, however, such perfect separability is rarely achievable and results in matrices with unequal entries on the antidiagonals. The last step of the basic SSA algorithm therefore performs a Hankelization of said matrices, i.e. a diagonal averaging is performed on all the $\mathbf{X}_i$ of Eq. (5) yielding matrices $\widetilde{\mathbf{X}}_i$ that have equal elements on the antidiagonals

$$\widetilde{\mathbf{X}} = \widetilde{\mathbf{X}}_\mathbf{I} + \widetilde{\mathbf{X}}_{\bar{I}} = \sum_{i \in I} \widetilde{\mathbf{X}}_i + \sum_{i \in \bar{I}} \widetilde{\mathbf{X}}_i \tag{6}$$

One can then e.g. easily reconstruct the approximation of the signal of interest through the eigentriples with indices I through the one-to-one correspondence between $\widetilde{\mathbf{X}}_\mathbf{I}$ and the respective time series $\widetilde{\mathbb{X}}_N = (\tilde{x}_1, \ldots, \tilde{x}_N)$ which provides an approximation of the entire time series $\mathbb{X}_N$ or some components of it, depending on the particular choice of indices I.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0056.v1

Peer-reviewed version available at Technologies 2018, 6, 26; doi:doi:10.3390/technologies6010026

The usefulness of basic SSA is illustrated in the example depicted in Figure 1 where the wandering baseline of an ECG signal (blue solid line) is removed by subtracting the trend reconstructed through SSA (with a window length of $M = 100$ and using the first two eigentriples $\widetilde{\mathbf{X}}_I = \widetilde{\mathbf{X}}_{i_1} + \widetilde{\mathbf{X}}_{i_2} \longrightarrow \widetilde{\mathbb{X}}_N = (\tilde{x}_1, \ldots, \tilde{x}_N)$ (red solid line)) from the original signal, i.e.

$$\mathbb{X}_{N_{\text{cleaned}}} = \mathbb{X}_N - \widetilde{\mathbb{X}}_N = (x_1, \ldots, x_N) - \left(\tilde{x}_{i_{11}}, \ldots, \tilde{x}_{i_{1N}}\right) - \left(\tilde{x}_{i_{21}}, \ldots, \tilde{x}_{i_{2N}}\right) \tag{7}$$

yielding the cleaned ECG signal (green solid line).

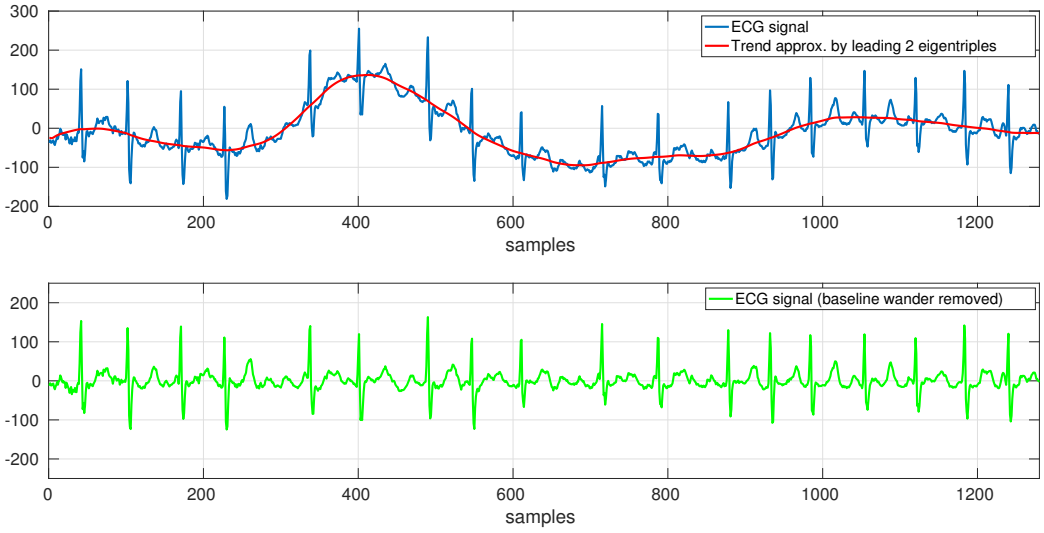

Figure 1. Application example of basic SSA: baseline wander removal. Showing excerpt 02:20 - 02:30 from record 14149m MIT-BIH ltdb [46]

For a more detailed discussion of SSA, we refer to two well-known monographs [37,48] in the field as well as [44,47] and references therein.

2.2. SSA Based Change Detection: Prior Art

The sequential application of SSA described in the following is based on work by Moskvina and Zhigljavsky and will be referred to as MZ in the remainder of this paper (*see* [35–37]). The need for an adaptation of basic SSA is due to the circumstance that it operates in batch mode and is therefore not suited for online change-point detection.

Assume a truly sequential problem in which observations $x_1, x_2, \ldots$ arrive one at a time. Having collected a sufficiently large number N of observations, MZ constructs the trajectory matrix $\mathbf{X}_B^{(n)}$ (the subindex B refers to 'base' for reasons that will become obvious in a moment) for time index n with $M \leq N/2, K = N - M + 1$ and performs the SVD and grouping steps as in basic SSA yielding an l-dimensional subspace $\mathcal{L}_I^{(n)} \subset \mathbb{R}^M$ spanned by the respective eigenvectors which captures the main structure of the signal.

The basic idea of MZ relies on the fact that the distance between the vectors $X_j^{(n)}, j = 1, \ldots, K$ and $\mathcal{L}_I^{(n)}$, controlled by the specific choice of I, can be reduced to rather small values. If monitoring of the series $\{x_t\}_{t=1}^N$ continues for $t > N$ without a change in the underlying data generating mechanism, the vectors $X_j, j > K$ are expected to remain relatively close to $\mathcal{L}_I^{(n)}$ while, on the other hand, if such a change were to occur at time $N + \tau$, the distance between $X_j, j \geq K + \tau$ and $\mathcal{L}_I$ would increase as it would move such vectors X_j out of the subspace $\mathcal{L}_I^{(n)}$ (*see* [36] at 2). Therefore, said distance can be used as a test statistic for change-point detection. Note that only the first three steps of basic SSA need to be performed since reconstruction of the original series is not required.

MZ constructs two matrices, the above mentioned *base matrix* $\mathbf{X}_B^{(n)}$, i.e. the trajectory matrix using data samples $x_{n+1},\ldots,x_{n+N}$, and a *test matrix* $\mathbf{X}_T^{(n)}$ using observations $x_{n+p+1},\ldots,x_{n+q+M-1}$. The former is subjected to SVD and used to obtain the subspace $\mathcal{L}_I^{(n)}$ while the latter serves to calculate the sum of squared Euclidean distances between its column vectors and $\mathcal{L}_I^{(n)}$. This process can be thought of as having two (possibly intersecting) windows (of M-dimensional data), of length K and $Q = q - p$ respectively, slide over the data.

Let N, M, l, p, q be fixed integers s.t. $l < M < N/2$ and $0 \leq p < q$. Then, for each $n = 0, 1, \ldots$ MZ proceeds as follows:

1. Apply SSA on the interval $[n+1, n+N]$ to get $\mathcal{L}_I^{(n)}$
   (a) Construct the trajectory/base matrix $\mathbf{X}_B^{(n)}$

$$\mathbf{X}_B^{(n)} = \begin{bmatrix} x_{n+1} & x_{n+2} & \cdots & x_{n+K} \\ x_{n+2} & x_{n+3} & \cdots & x_{n+K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n+M} & x_{n+M+1} & \cdots & x_{n+N} \end{bmatrix} \tag{8}$$

   where $K = N - M + 1$.

   (b) Singular Value Decomposition of $\mathbf{X}_B^{(n)}$.

   (c) Selection of $I = \{i_1, \ldots, i_l\}, l < d \leq M$ with $d = \max(i, \text{ such that } \lambda_i > 0)$.

2. Construct test matrix $\mathbf{X}_T^{(n)}$ on the interval $[n+p+1, n+q+M-1]$

$$\mathbf{X}_T^{(n)} = \begin{bmatrix} x_{n+p+1} & x_{n+p+2} & \cdots & x_{n+q} \\ x_{n+p+2} & x_{n+p+3} & \cdots & x_{n+q+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n+p+M} & x_{n+p+M+1} & \cdots & x_{n+q+M-1} \end{bmatrix} \tag{9}$$

3. Compute the detection statistic $D_{n,I,p,q}$

$$D_{n,I,p,q} = \sum_{j=p+1}^{q} \left[\left(X_j^{(n)}\right)^T X_j^{(n)} - \left(X_j^{(n)}\right)^T \mathbf{U}_I^{(n)} \left(\mathbf{U}_I^{(n)}\right)^T X_j^{(n)} \right] \tag{10}$$

where $X_j^{(n)} = [x_{n+j}, \ldots, x_{n+j+M-1}]^T$ and $\mathbf{U}_I^{(n)} = \left[U_{i_1}^{(n)}, \ldots, U_{i_l}^{(n)}\right]$ is the $M \times l$ matrix of eigenvectors spanning $\mathcal{L}_I^{(n)}$, i.e. $D_{n,I,p,q}$ is the sum of squared Euclidean distances between the columns of $\mathbf{X}_T^{(n)}$ and $\mathcal{L}_I^{(n)}$. MZ normalizes the sum of distances $D_{n,I,p,q}$ to the number of elements in $\mathbf{X}_T^{(n)}$

$$\tilde{D}_{n,I,p,q} = \frac{D_{n,I,p,q}}{MQ} \tag{11}$$

and further normalizes the test statistic as

$$S_{n,I,p,q} = \frac{\tilde{D}_{n,I,p,q}}{v_n} \tag{12}$$

such that it does not depend on the unknown variance of the noise (*see* [35] at 28) with v_n being an estimator of $\tilde{D}_{n,I,p,q}$, e.g. $v_n = \tilde{D}_{m,I,0,K}$ with $m \leq n$ such that the hypothesis of no change can be accepted.

4. Monitoring of $S_{n,I,p,q}$ using CUSUM-type Control Charts

MZ then constructs the following CUSUM-type control chart

$$W_1 = S_{1,I,p,q}, \quad W_{n+1} = \max\left(0, W_n + S_{n+1,I,p,q} - S_{n,I,p,q} - \kappa\right), \quad n \geq 1 \tag{13}$$

with κ suggested as $\kappa = 1/\left(3\sqrt{MQ}\right)$ (*see* [35] at 29) and threshold $h_{MZ} = 1 + 1.9\sqrt{M}$ (*see* [35] at 35). A change-point at n is then declared if

$$W_n \geq h_{MV} \tag{14}$$

holds.

3. The Proposed Method (l-SSA-CPD)

While MZ provides a powerful methodology that could be applied directly to raw ECG (or PPG) data, it exhibits some drawbacks (for the particular application at hand) that motivated the development of the novel approach to be presented below which we shall refer to as *lightweight-SSA-ChangePointDetection* (l-SSA-CPD).

3.1. Motivation and Informal Description of the Improvements

As discussed in the preceding Section, MZ makes use of two (possibly intersecting) windows that are slid over the observed time series, one comprising the data that is embedded to form the trajectory matrix, which is then decomposed by means of SVD to identify an appropriate low(er)-dimensional subspace, and another one containing new (or, in case of overlap, a combination of old and new) observations whose distance to said low-dimensional subspace is then used as a test statistic. This entails the quite burdensome step of performing a SVD every time a new data sample becomes available.

We shall first highlight the main improvements of our method prior to its formal description.

- Low Computational Complexity

  Small variations over time are intrinsic to cardiac signals and may, besides noise and motion artifacts, e.g. be due to *Heart Rate Variability*. Contrary to anomalies caused by abnormal cardiac excitation phenomena, these changes in the time between consecutive R-peaks are subtle and often not readily discernible. Most importantly, they do not induce changes as severe as to change the signal's main characteristics which are captured through the decomposition and grouping stages of SSA. This is illustrated in Figure 2 which shows a raw (unfiltered) ECG signal with two distinctly shaped PVCs (highlighted in purple) in the third quarter of the excerpt.

  For the task at hand, performing the SVD of a newly generated trajectory matrix each time a new data point becomes available is not strictly necessary. We are able to drastically reduce the computational burden by generating only one initial trajectory matrix $\mathbf{X}_B^{(0)}$ and relying on the obtained reference subspace $\mathcal{L}_I^{(0)}$ throughout the monitoring. This is shown in Figure 2 with the section of the signal highlighted through gray and blue backgrounds representing the intervals used to generate $\mathbf{X}_B^{(n)}$ and $\mathbf{X}_T^{(n)}$, respectively. Note how in MZ (left part of Figure 2) both $\mathbf{X}_B^{(n)}$ and $\mathbf{X}_T^{(n)}$ are being slid over the observations while in our algorithm (right part of Figure 2) only $\mathbf{X}_T^{(n)}$ is a sliding window since $\mathbf{X}_B^{(n)}|_{\forall n} = \mathbf{X}_B^{(0)}$.
- Simplicity

  By sliding only a single instead of two windows over the time series the entire procedure is simplified and benefits from a reduction in tuning parameters.

  In fact, while the total number Q of columns in $\mathbf{X}_T$ is of course relevant, p and q are not since, due to $\mathbf{X}_B^{(n)}|_{\forall n} = \mathbf{X}_B^{(0)}$ an overlap of $\mathbf{X}_B$ and $\mathbf{X}_T$ can only occur in the first $N - p$ samples for $p < N$. While our algorithm allows for such an initial overlap of $\mathbf{X}_B$ and $\mathbf{X}_T$, the following discussion is purposely limited to $p = N < q$, which is in line with recommendations by Moskvina and Zhigljavsky who point out that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0056.v1

Peer-reviewed version available at Technologies 2018, 6, 26; doi:doi:10.3390/technologies6010026

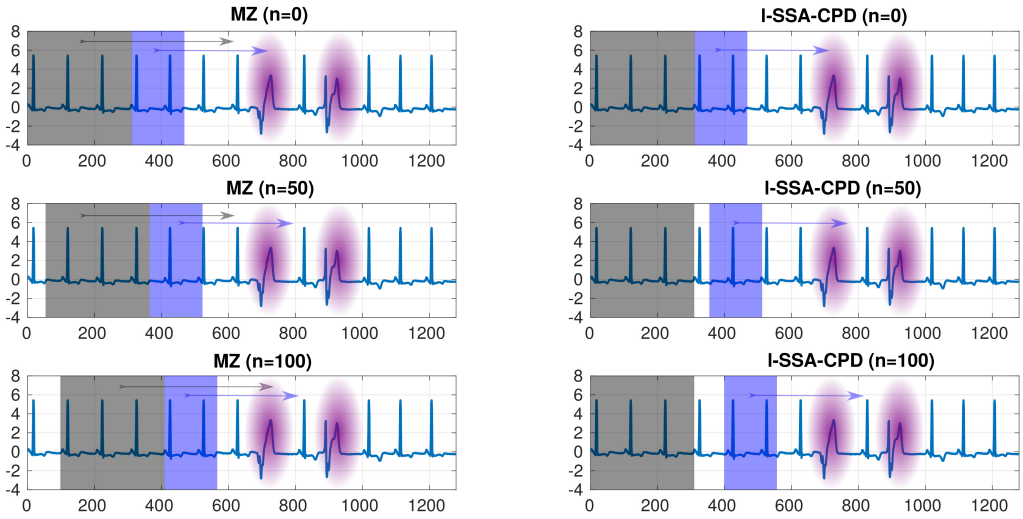

Figure 2. Comparison of MZ (left) and l-SSA-CPD (right). The computational burden of l-SSA-CPD is greatly reduced compared to MZ by relying on the reference subspace obtained from an initial, non-sliding trajectory matrix (illustrated by the gray background area). Furthermore, note that l-SSA-CPD's reference subspace remains locked on the main signal's characteristics while in MZ, since the reference subspace is updated at each observation (illustrated by the gray area sliding as well), it will lock on the two anomalies (highlighted in pink) for some time as the two moving windows are pass over them.

$p = N, q = N + 1$ and accordingly $Q = 1$ is a very reasonable choice if minimizing the detection delay is of importance, since $Q > 1$ entails a smoother behavior of the test statistic and thus a loss of agility (*see, e.g.* [35] at 30).

The question as to whether or not $\mathbf{X}_B$ and $\mathbf{X}_T$ should overlap and if so by how much is therefore removed. Furthermore, since $p = N, q = N + 1$ can generally be recommended (*see* Section 4), we can omit both tuning parameters p and q.

- Augmentation of Test Statistic by considering the angle between $\mathcal{L}_I^{(0)}$ and $X_j^{(n)}$

Some authors [49–51] successfully proposed a modified version of MZ, wherein the test statistic is based on angles rather than on Euclidean distances. While both approaches are viable on their own merits, we chose to merge them as they augment each other yielding a test statistic that, according to our results on raw ECG and PPG records, performs favorably compared to the test statistic constructed using either one on its own.

In other words, we augment and improve upon the test statistic of MZ by making use of the information from the angles between $\mathcal{L}_I^{(0)}$ and $X_j^{(n)}$ as well.

- Improved thresholding through Sequential Ranks CUSUM

MZ provides further potential for improvement by employing a CUSUM-type control chart (*see* Eq. (13-14)) whose control limit (or threshold) h is obtained through suitable normal approximation and asymptotic considerations (*see* [35] at 31; *see also* [36] at 8). The nuisance of having to properly normalize the test statistic (*see* Eq. (12)) is a direct consequence of this design choice.

We instead propose the use of McDonald's Sequential Ranks CUSUM (SRC) [52] which we deem to be more appropriate and in line with the model-free nature of SSA.

- Restarting of SRC control chart after it signaled

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0056.v1

Peer-reviewed version available at Technologies 2018, 6, 26; doi:doi:10.3390/technologies6010026

Lastly, to allow for the detection of multiple and potentially nearby change-points we restart the SRC every time after it signaled an anomaly by exceeding the preset threshold h_{SRC}.

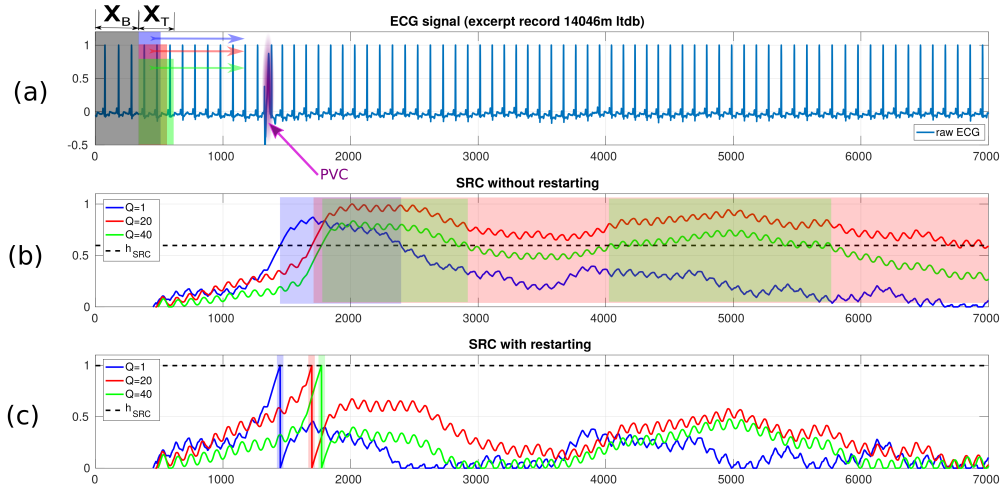

Figure 3. Restarting the SRC each time the threshold was hit (c) is indicated to avoid potentially lengthy delays before the anomaly propagates through and clears $\mathbf{X}_T$ (b)

This is indicated since otherwise, depending on the extent of the anomaly (in terms of number of samples) and how it relates to the embedding dimension M as well as the number Q of columns in the Hankel matrix $\mathbf{X}_T$, it may take some time before the anomaly propagates through and clears $\mathbf{X}_T$ (i.e. our sliding window) thereby resulting in a return of the test statistic to its 'baseline level'. This is illustrated in Figure 3 where, as depicted in (a), three windows (differing in size) with $Q = \{1, 20, 40\}$ are slid over an ECG containing a single PVC (highlighted in purple and marked with an arrow). As can be seen in (b), while $Q = 1$ is a feasible choice even without restarting the control chart after exceeding the threshold h_{SRC}, since the respective SRC returns to values below h_{SRC} after a relatively long but perhaps still acceptable amount of time, the same cannot be said for $Q = \{20, 40\}$. Part (c) of Figure 3 illustrates the clear benefits of restarting the control charts, in that regardless of the choice of Q the PVC is detected and monitoring for further changes can swiftly resume. As was to be expected, $Q = 1$ is favorable in terms of detection delay.

3.2. Formal Description of l-SSA-CPD

To allow for better comparison, we use the notation introduced in Section 2.2 as far as possible.
Let N, M, l, p, q be fixed integers s.t. $l < M < N/2$ and $0 \leq p < q$. Then our method proceeds as follows:

1. Initialization at $n = 0$

   SSA is applied on the interval $[n+1, n+N]$ to get $\mathcal{L}_I = \mathcal{L}_I^{(n=0)}$, which akin to MZ involves

   (a) Construction of the trajectory/base matrix $\mathbf{X}_B = \mathbf{X}_B^{(0)} = \mathbf{X}_B^{(n=0)}$.

$$\mathbf{X}_B = \begin{bmatrix} x_{n+1} & x_{n+2} & \dots & x_{n+K} \\ x_{n+2} & x_{n+3} & \dots & x_{n+K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n+M} & x_{n+M+1} & \dots & x_{n+N} \end{bmatrix} \tag{15}$$

   where $K = N - M + 1$.

(b) Singular Value Decomposition of $\mathbf{X}_B$.

(c) Selection of $I = \{i_1, \ldots, i_l\}, l < d \leq M$ with $d = \max(i, \text{ such that } \lambda_i > 0)$.

Then, for each $n = 0, 1, \ldots$ we proceed as follows

2. Construct test matrix $\mathbf{X}_T^{(n)}$ on the interval $[n+p+1, n+q+M-1]$

$$\mathbf{X}_T^{(n)} = \left[X_{j|_{j=p+1,\ldots,q}}^{(n)}\right] = \begin{bmatrix} x_{n+p+1} & x_{n+p+2} & \cdots & x_{n+q} \\ x_{n+p+2} & x_{n+p+3} & \cdots & x_{n+q+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n+p+M} & x_{n+p+M+1} & \cdots & x_{n+q+M-1} \end{bmatrix} \tag{16}$$

with $X_j^{(n)} = [x_{n+j}, \ldots, x_{n+j+M-1}]^T$.

3. Compute the detection statistics $D_{n,I,p,q}^{\dagger_{1,\ldots,3}}$

$$D_{n,I,p,q}^{\dagger_1} = \frac{1}{Q} \sum_{j=p+1}^{q} \left[\left(X_j^{(n)}\right)^T X_j^{(n)} - \left(X_j^{(n)}\right)^T \mathbf{U}_I \left(\mathbf{U}_I\right)^T X_j^{(n)} \right] \tag{17}$$

$$D_{n,I,p,q}^{\dagger_2} = 1 - \cos\left[\angle\left(\mathbf{X}_T^{(n)}, \mathcal{L}_I\right)\right] \tag{18}$$

$$D_{n,I,p,q}^{\dagger_3} = D_{n,I,p,q}^{\dagger_1} \circ D_{n,I,p,q}^{\dagger_2} \tag{19}$$

with

$$\angle\left(\mathbf{X}_T^{(n)}, \mathcal{L}_I\right) = \angle\left(\mathbf{X}_T^{(n)}, \mathbf{U}_I\right) = \frac{1}{Q} \sum_{j=p+1}^{q} \frac{1}{l} \left[\sum_{k=i_1}^{i_l} \arccos\left(\frac{\langle X_j^{(n)}, U_k \rangle}{\left\| X_j^{(n)} \right\| \|U_k\|} \right) \right] \tag{20}$$

taking values in $\left[0, \frac{\pi}{2}\right]$, accordingly $D_{n,I,p,q}^{\dagger_2} \in [0,1]$, $\mathbf{U}_I = [U_{i_1}, \ldots, U_{i_l}]$ being the $M \times l$ matrix of eigenvectors spanning $\mathcal{L}_I$, and $\circ$ denoting the Hadamard (element-wise) product.

4. Monitoring of $D_{n,I,p,q}^{\dagger_{1,\ldots,3}}$ using the Sequential Ranks CUSUM Control Chart

Let us denote the sequential rank of $D_{n,I,p,q}^{\dagger_{1,\ldots,3}}$ as

$$R_n = 1 + \sum_{r=1}^{n-1} \max\left(0, D_{n,I,p,q}^{\dagger_{1,\ldots,3}} - D_{r,I,p,q}^{\dagger_{1,\ldots,3}}\right) \tag{21}$$

The Sequential Ranks CUSUM is then

$$C_n = \max\left(0, C_{n-1} + \frac{R_n}{n+1} - k_{\text{SRC}}\right), \quad n \geq 1 \tag{22}$$

with $C_0 = 0$ and k_{SRC} being a reference constant.

The SRC then signals and a change-point at n is declared if

$$C_n \geq h_{\text{SRC}} \tag{23}$$

holds, i.e. if C_n exceeds a predetermined control limit h_{SRC}.

It can be shown [52] that, given that no change in the monitored signal occurred, the quantities $\frac{R_n}{n+1}$ are independent and discrete uniform on

$$\{\frac{1}{n+1}, \frac{2}{n+1}, \cdots, \frac{n}{n+1}\}$$

which represents a crucial advantage of the SRC in that it implies that for any k_{SRC} we can obtain the control limit h_{SRC} without the need for any historical training data or further assumptions through simulations as follows:

Algorithm 1: Calculate h_{SRC} for fixed k_{SRC}

(a) Set a constant N_{SRC}

(b) **for** $i = 1$ **to** B **do**

i. Construct the set of random variables $\{Y_n\}_{n=1}^{N_{\mathrm{SRC}}}$ as discrete uniform on $\{\frac{1}{n+1}, \frac{2}{n+1}, \cdots, \frac{n}{n+1}\}$

ii. Construct $C_{\mathrm{SRC}_n} = \max\{0, C_{\mathrm{SRC}_{n-1}} + Y_n - k_{\mathrm{SRC}}\}$

iii. Extract the maximum value of $\{C_{\mathrm{SRC}_n}\}_{n=1}^{N_{\mathrm{SRC}}}$

end

(c) Set the control limit h_{SRC} as the $B \cdot (1 - ARL_0{}^{-1})$ ordered extracted maximum value with ARL_0 being the nominal in-control average run length (ARL) (*see* Appendix A).

4. Performance Evaluation

To evaluate and assess the performance and utility of our method we use records which are publicly available through Physiobank [46], a vast and commonly used resource for ECG and other biophysiological data. In particular, since we claim our method to be suitable for both ECG and PPG data, we found the Physionet Challenge 2015 training database (PC15) [45] to be of particular interest as it provides a collection of synchronized ECG and PPG recordings from which we chose a subset similar to the one used in [53]. With PC15 records not being annotated, we purposely chose to limit our evaluation to records containing PVCs (with different frequencies of occurrence) since they can quite accurately be spotted by careful visual inspection.

Before presenting some results, it seems appropriate to briefly restate the goal of our method, which is to provide a lightweight, model-free tool capable of providing a rough assessment under tight constraints on computational resources, e.g. as a pre-screening tool. It is therefore not to substitute for but rather to complement more sophisticated and (computationally) expensive procedures.

4.1. Performance Metrics

In reporting our results we rely on established metrics commonly used in the literature and report *sensitivity* (Se), *specificity* (Sp), and *accuracy* (Acc) defined as

$$Se = \frac{TP}{TP + FN} \tag{24}$$

$$Sp = \frac{TN}{TN + FP} \tag{25}$$

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{26}$$

with TP, FP, TN, FN being the number of *true positives*, *false positives*, *true negatives*, and *false negatives*, respectively.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0056.v1

Peer-reviewed version available at Technologies 2018, 6, 26; doi:doi:10.3390/technologies6010026

Accordingly, *sensitivity* quantifies the ability to correctly detect actual anomalies while vice versa *specificity* quantifies the proportion of non-abnormal segments that are correctly identified as such. *Accuracy*, on the other hand, assesses the overall performance in terms of both correctly identified abnormal and non-abnormal segments.

Note that there is a nonzero detection delay introduced by the use of a control chart. Typically, said delay tends to be longer for nonparametric control charts such as the SRC compared to parametric charts (*see, e.g.* [52]). Use of the latter however would require imposing a parametric model and therefore inevitably conflict with our goal of minimizing (distributional) assumptions as much as possible.

For an event occurring at time instance n we allow for a certain detection delay τ_d and consider a signal from the control chart as true positive if it falls in the interval $[n, n+\tau_d]$. All results presented here were obtained using $\tau_d = N$, i.e. we allow for a detection delay less or equal to the length of the interval used to construct the initial trajectory matrix $\mathbf{X_B}$.

Furthermore, note that when directly comparing (synchronized) ECG and PPG signals there is an inherent delay (between the R-peak of the ECG and the respective pulse peak in the PPG) in the PPG signal due to the propagation delay of the pulse pressure wave through the arterial system. This is commonly referred to as *Pulse Arrival Time* (PAT) or *Pulse Transit Time* (PTT) (*see, e.g.* [9,54]) and illustrated on a short data excerpt (containing a single PVC, highlighted by blue and red backgrounds for ECG and PPG, respectively) in Figure 4. To account for the PTT delay, when dealing with PPG signals we shift the interval $[n, n+\tau_d]$ by $\lfloor 2M/5 \rfloor$, taken to be a rough estimate of the actual PTT.

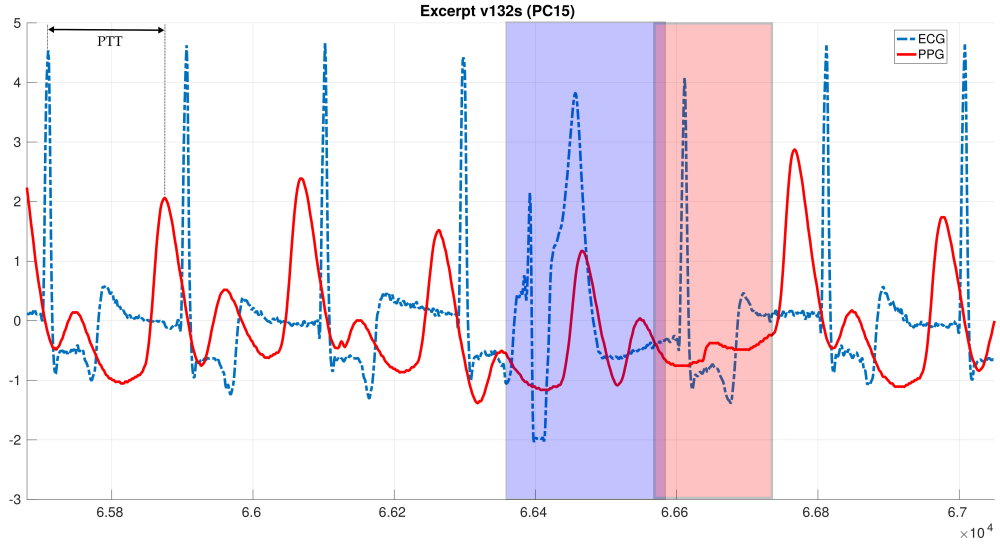

Figure 4. Excerpt of a synchronized recording of ECG (blue dash-dotted) and PPG (red solid) containing a single ectopic beat, highlighted with blue and red backgrounds respectively. Note the shift between the R-peak of the ECG and the respective pulse peak of the PPG, known as Pulse Transit Time (PTT)

4.2. Setup and l-SSA-CPD Parameters

The PC15 database [45] provides a collection of ECG and PPG recordings from which we chose a subset similar to the one used in [53]. With PC15 records not being annotated, we purposely chose to limit our evaluation to records containing PVCs (with different frequencies of occurrence), since they can quite accurately be spotted by careful visual inspection, eventually including 8 records (composed of two ECG and one PPG signal per record) with varying frequency of PVC occurrence in our analysis. The records are approximately 5 minutes long with the sampling frequency being 250Hz yielding about 75000 observations each.

Since an in-depth discussion of how to select important SSA tuning parameters, most notably window length M and number of eigentriples used (i.e. selection of I), would be beyond the scope of this paper (*see, e.g.*

[35–37,44,47,48] and references therein), it shall suffice to briefly discuss our settings and the rationale behind them.

Consider a periodic signal with period T, then for SSA to capture the main structure of the signal it is important that M be at least equal to T. Taking into account the physiological limits on HR and the sampling frequency of our signals, $M = 300$ appears to be a safe and reasonable choice. Accordingly, since as discussed in Section 2.1.1 we impose $M \leq N/2$, we set $N = 2M = 600$. Furthermore, we set I to contain the leading l eigentriples such as to account for 90% of the data's variance. As for the SRC's control limit, we use $h_{\mathrm{SRC}} = 79.4107$ which we obtained through Algorithm 1 for $B = 10^6, N_{\mathrm{SRC}} = 5000, ARL_0 = 5000, k_{\mathrm{SRC}} = 0.5$.

It shall further be emphasized that we apply l-SSA-CPD on the raw unfiltered data without any preprocessing steps. Clearly, suitable preprocessing steps might further enhance performance, the objective here however is to ascertain whether or not usable pre-screening information (pertaining to presence or absence of anomalies) can be obtained by solely applying our l-SSA-CPD with very general parameter settings. A direct performance comparison to MZ is therefore omitted for two main reasons:

- MZ would be computationally prohibitively expensive.
  Recall that the PC15 records are approximately $75 \cdot 10^3$ samples long, requiring computing the SVD of a 300×301 trajectory matrix, assuming $M = 300, N = 600, K = N - M + 1, Q = 1$ about 74100 times as opposed to just once for l-SSA-CPD (*see* Figure 2).
- We aim to assess whether, based on its own merits, the performance of l-SSA-CPD suffices to be considered for potential real life applications such as the use case presented in this paper. To further this goal an in-depth comparative analysis to competing algorithms is not required and deemed to be beyond the scope of this paper.

4.3. Performance Evaluation Using ECG Signals

Table 1 shows the experimental results obtained by applying l-SSA-CP configured as described above to 8 PC15 ECG records with varying length $Q = \{1, 5, 10\}$ of the test matrix $\mathbf{X}_T^{(n)}$.

A crucial condition for l-SSA-CPD to work properly is that the first N samples, which are embedded to form the trajectory matrix $\mathbf{X_B}$, be an adequate representation of the underlying signal. In other words, we require this initial segment to be free of anomalies. If an anomaly occurs in the first N samples, those samples are to be discarded. This was the case for record t662s, which contains a premature ventricular contraction at about $n = 332 < N$ and required us to discard the first few hundred observations.

Examining the entries of Table 1 it is apparent that l-SSA-CPD performs well, especially keeping in mind that in our setup it is applied with fairly general parameters to raw, unfiltered ECG traces. The bottom of the table highlighted in red presents the average performance over the entire 8 records for the three different test statistics $\{D^{\dagger 1}_{n,I,p,q}, D^{\dagger 2}_{n,I,p,q}, D^{\dagger 3}_{n,I,p,q}\}$ and test matrix widths $Q = \{1, 5, 10\}$.

Note the increased performance of l-SSA-CPD with test statistic $D^{\dagger 3}_{n,I,p,q}$ and the rather small benefit (if any) of using larger values for Q. These findings corroborate our recommendations made in Section 3.1 to use $D^{\dagger 3}_{n,I,p,q}$ and $Q = 1$, with the latter being in agreement with results reported by other authors (*see, e.g.* [35,36]). Results obtained using the second ECG trace of the 8 PC15 records were similar and are therefore omitted.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0056.v1

Peer-reviewed version available at Technologies 2018, 6, 26; doi:doi:10.3390/technologies6010026

Table 1. Detection on PC15 ECG trace I

Record	Q	1			5			10		
	Statistic	$D^{\dagger_1}_{n,I,p,q}$	$D^{\dagger_2}_{n,I,p,q}$	$D^{\dagger_3}_{n,I,p,q}$	$D^{\dagger_1}_{n,I,p,q}$	$D^{\dagger_2}_{n,I,p,q}$	$D^{\dagger_3}_{n,I,p,q}$	$D^{\dagger_1}_{n,I,p,q}$	$D^{\dagger_2}_{n,I,p,q}$	$D^{\dagger_3}_{n,I,p,q}$
v132s	**Se**	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Sp	0.9998	0.9997	0.9997	0.9998	0.9997	0.9997	0.9998	0.9997	0.9997
	Acc	0.9998	0.9997	0.9997	0.9998	0.9997	0.9997	0.9998	0.9997	0.9997
v253l	**Se**	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Sp	0.9995	0.9996	0.9996	0.9995	0.9996	0.9996	0.9995	0.9996	0.9996
	Acc	0.9995	0.9996	0.9996	0.9995	0.9996	0.9996	0.9995	0.9996	0.9996
v255l	**Se**	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Sp	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995
	Acc	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995
v368s	**Se**	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Sp	0.9992	0.9990	0.9992	0.9992	0.9990	0.9992	0.9992	0.9992	0.9991
	Acc	0.9992	0.9990	0.9992	0.9992	0.9990	0.9992	0.9992	0.9990	0.9991
v557l	**Se**	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000
	Sp	0.9996	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995
	Acc	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995
t662s	**Se**	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Sp	0.9994	0.9993	0.9993	0.9994	0.9993	0.9993	0.9994	0.9992	0.9993
	Acc	0.9994	0.9993	0.9993	0.9994	0.9993	0.9993	0.9994	0.9992	0.9993
a746s	**Se**	0.6667	0.6667	0.8333	0.6667	0.6667	0.8333	0.6667	0.6667	0.8333
	Sp	0.9980	0.9981	0.9980	0.9980	0.9981	0.9980	0.9980	0.9981	0.9980
	Acc	0.9980	0.9981	0.9980	0.9980	0.9981	0.9980	0.9980	0.9981	0.9980
v831l	**Se**	0.4583	0.4167	0.5000	0.4583	0.4167	0.5417	0.4583	0.5000	0.5417
	Sp	0.9994	0.9993	0.9993	0.9994	0.9992	0.9993	0.9994	0.9993	0.9993
	Acc	0.9992	0.9991	0.9992	0.9992	0.9991	0.9992	0.9992	0.9991	0.9992
$\bar{\Sigma}_{\text{ECG}}$	**Se**	0.7656	0.8854	0.9167	0.7656	0.8854	0.9219	0.7656	0.8958	0.9219
	Sp	0.9993	0.9992	0.9993	0.9993	0.9992	0.9993	0.9993	0.9993	0.9992
	Acc	0.9993	0.9992	0.9993	0.9993	0.9992	0.9993	0.9993	0.9992	0.9992

4.4. Performance Evaluation Using PPG Signals

Experimental results obtained by applying l-SSA-CPD to the PPG trace of the same 8 PC15 records are shown in Table 2, again with varying length $Q = \{1, 5, 10\}$ of the test matrix $\mathbf{X}_T^{(n)}$.

Comparing the entries of Table 2 with those Table 1 we notice an overall drop in performance. Nevertheless, l-SSA-CPD still manages to provide reasonable results. Furthermore, the recommendation of using $D^{\dagger_3}_{n,I,p,q}$ and $Q = 1$ is shown to hold for the PPG traces as well.

It should be pointed out that, as has already been observed by other authors (*see, e.g.* [53]), there are some inconsistencies in the PC15 records in that the some of the supposedly synchronized PPG traces exhibit an unusual delay not consistent with the assumption that the PPG pulse peak should have an offset equal to the PTT with respect to the respective R peak in the ECG. Pflugradt et al. attribute this occasional unusual offset to glitches in the original measurement setup (*see* [53] at 11) and we assume it to have negatively impacted the performance characteristics presented in Table 2 since we allow only for a very limited detection delay τ_d.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 December 2017 doi:10.20944/preprints201712.0056.v1

Peer-reviewed version available at Technologies 2018, 6, 26; doi:doi:10.3390/technologies6010026

Table 2. Detection on PC15 PPG trace

Record	Q	1			5			10		
	Statistic	$D^{\dagger_1}_{n,I,p,q}$	$D^{\dagger_2}_{n,I,p,q}$	$D^{\dagger_3}_{n,I,p,q}$	$D^{\dagger_1}_{n,I,p,q}$	$D^{\dagger_2}_{n,I,p,q}$	$D^{\dagger_3}_{n,I,p,q}$	$D^{\dagger_1}_{n,I,p,q}$	$D^{\dagger_2}_{n,I,p,q}$	$D^{\dagger_3}_{n,I,p,q}$
v132s	**Se**	0.7097	0.9677	0.8387	0.7097	0.9677	0.8387	0.6774	0.9677	0.8387
	Sp	0.9995	0.9993	0.9994	0.9995	0.9993	0.9994	0.9995	0.9993	0.9994
	Acc	0.9994	0.9993	0.9993	0.9994	0.9993	0.9993	0.9993	0.9993	0.9993
v253l	**Se**	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Sp	0.9996	0.9997	0.9997	0.9996	0.9997	0.9996	0.9996	0.9997	0.9997
	Acc	0.9996	0.9997	0.9997	0.9996	0.9997	0.9996	0.9996	0.9997	0.9997
v255l	**Se**	0.8594	0.8594	0.8438	0.8594	0.8594	0.8438	0.8594	0.8750	0.8438
	Sp	0.9999	0.9998	0.9998	0.9999	0.9998	0.9998	0.9999	0.9998	0.9998
	Acc	0.9998	0.9997	0.9997	0.9998	0.9997	0.9997	0.9998	0.9997	0.9997
v368s	**Se**	0.2857	0.2857	0.4286	0.2857	0.4286	0.4286	0.2857	0.4286	0.4286
	Sp	0.9983	0.9986	0.9985	0.9983	0.9986	0.9985	0.9983	0.9986	0.9985
	Acc	0.9982	0.9986	0.9984	0.9982	0.9986	0.9984	0.9982	0.9986	0.9984
v557l	**Se**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Sp	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997
	Acc	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996
t662s	**Se**	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	Sp	0.9988	0.9988	0.9988	0.9988	0.9988	0.9988	0.9988	0.9988	0.9988
	Acc	0.9988	0.9988	0.9988	0.9988	0.9988	0.9988	0.9988	0.9988	0.9988
a746s	**Se**	0.8333	1.0000	1.0000	0.8333	1.0000	1.0000	0.8333	1.0000	0.8333
	Sp	0.9995	0.9996	0.9996	0.9995	0.9996	0.9995	0.9995	0.9996	0.9995
	Acc	0.9995	0.9996	0.9996	0.9995	0.9996	0.9995	0.9995	0.9996	0.9995
v831l	**Se**	0.2083	0.3333	0.3333	0.2083	0.3333	0.3750	0.2833	0.3333	0.3750
	Sp	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997
	Acc	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995	0.9995
$\bar{\Sigma}_{\text{PPG}}$	**Se**	0.6121	0.6808	0.6806	0.6120	0.6986	0.6858	0.6174	0.7006	0.6649
	Sp	0.9994	0.9994	0.9994	0.9994	0.9994	0.9994	0.9994	0.9994	0.9994
	Acc	0.9993	0.9994	0.9993	0.9993	0.9994	0.9993	0.9993	0.9994	0.9993

5. Discussion and Outlook

In this paper, we have proposed e a novel lightweight and model-free approach for the online detection of cardiac anomalies such as ectopic beats in ECG or PPG signals based on the change detection capabilities of Singular Spectrum Analysis (SSA) and nonparametric rank-based cumulative sum (CUSUM) control charts. The procedure is able to quickly detect anomalies without requiring the identification of fiducial points such as R-peaks and is computationally significantly less demanding than previously proposed SSA-based approaches. This is accomplished by modifying the conventional SSA-based change detection algorithm such that the computationally expensive task of computing the SVD is only performed at the very beginning instead of each time a new data point becomes available. Furthermore, our procedure uses more elaborate test statistics that take into account the information derived from the angle between data vectors representing new observations and the subspace representing the signals characteristics as well as the euclidean distances. Lastly, our procedure differs from previous approaches also in that rank-based control limits and the reinitialization of control charts after an anomaly has been detected are used.

Using a set of ECG and PPG records we demonstrated that the direct application of our l-SSA-CPD without any further pre- or post-processing yields not only viable but surprisingly accurate results with an average sensitivity and specificity of 0.9167 and 0.9993 for ECG and 0.6806 and 0.9994 for PPG records, respectively.

With regards to the selection and fine-tuning of SSA-parameters, the performance evaluation on records containing cardiac anomalies other than PVCs and a comparative performance analysis, questions for future works are left open.

Acknowledgments: The work of M. Lang was supported by the 'Excellence Initiative' of the German Federal and State Governments and the Graduate School of Excellence Computational Engineering at Technische Universtät Darmstadt. The

views expressed in this article are solely those of the author in his private capacity and do not necessarily reflect the views of Technische Universtät Darmstadt or any other organization.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
ARL	Average Run Length
AV	Atrioventricular Node
CUSUM	Cumulative Sum Control Chart
ECG	Electrocardiography
HR	Heart Rate
PAC	Premature Atrial Contraction
PAT	Pulse Arrival Time
PCA	Principal Component Analysis
PPG	Photoplethysmography
PR	Pulse Rate
PTT	Pulse Transit Time
PVC	Premature Ventricular Contraction
SA	Sinoatrial Node
SRC	Sequential Ranks CUSUM Control Chart
SSA	Singular Spectrum Analysis
SVD	Singular Value Decomposition

Appendix A. CUSUM Control Charts

Consider an observed sequence $\{x(n), n \geq 1\}$ of independent random variables such that $\{x(1), \ldots, x(\tau-1)\} \sim F$ and $\{x(\tau), x(\tau+1), \ldots\} \sim G$, i.e. a distributional shift $F \rightarrow G$ occurs at time instance τ.

Under the assumption that F and G were normally distributed with known parameters, Page's CUSUM [55] represents the gold-standard change detection technique and can be computed sequentially as

$$C_{\mathrm{C}}(0) = 0, \quad C_{\mathrm{C}}(n) = \max\{0, C_{\mathrm{C}}(n-1) + x(n) - k_{\mathrm{C}}\}, \quad n \geq 1 \tag{A1}$$

The CUSUM signals, thereby declaring a distributional shift to have occurred, if

$$C_{\mathrm{C}}(n) > h_{\mathrm{C}} \tag{A2}$$

with pre-specified control limit/threshold and reference constant h_{C} and k_{C}, respectively. h_{C} and k_{C} are chosen such that a nominal average run length (ARL) of ARL_0 is attained when the control chart operates in-control, i.e. without distributional shifts occurring (*see, e.g.* [56]).

Formally, the in-control ARL is defined as the expected time until a change is signaled under F, i.e.

$$ARL = E_F \inf\{n > 0 : C_{\mathrm{C}}(n) > h_{\mathrm{C}}\}. \tag{A3}$$

which can be interpreted as akin to setting a nominal type-I error level in hypothesis testing with the closeness of the actual in-control ARL to ARL_0 then being an indicator of the CUSUM chart's robustness [57]. It is well known that, under some regularity conditions, choosing $k_{\mathrm{C}} = \delta/2$, with δ being the shift in the transition $F \rightarrow G$, is optimal [58].

References

1. Mukhopadhyay, S.C., Ed. Wearable Electronics Sensors - For Safe and Healthy Living, 1st ed.; Springer, 2015.

2. Selke, S., Ed. Lifelogging: Digital self-tracking and Lifelogging - between disruptive technology and cultural transformation, 1st ed.; Springer, 2016.
3. Iaizzo, P.A., Ed. Handbook of Cardiac Anatomy, Physiology, and Devices, 3rd ed.; Springer International Publishing, 2015.
4. Gupta, R.; Mitra, M.; Bera, J. ECG Acquisition and Automated Remote Processing, 1st ed.; Springer, 2014.
5. Gacek, A.; Pedrycz, W. ECG Signal Processing, Classification and Interpretation: A Comprehensive Framework of Computational Intelligence, 1st ed.; Springer, 2012.
6. Kiasaleh, K. Biological Signals Classification and Analysis, 1st ed.; Springer, 2015.
7. Webster, J., Ed. Design of Pulse Oximeters, 1st ed.; CRC Press, 1997.
8. Sazonov, E.; Neuman, M.R., Eds. Wearable Sensors: Fundamentals, Implementation and Applications, 1st ed.; Academic Press, 2014.
9. Allen, J. Photoplethysmography and its application in clinical physiological measurement. Physiological Measurement **2007**, 28, R1–R39.
10. Adibi, S., Ed. Mobile Health - A Technology Road Map, 1st ed.; Springer International Publishing, 2015.
11. Holzinger, A.; Röcker, C.; Ziefle, M., Eds. Smart Health - Open Problems and Future Challenges, 1st ed.; Springer International Publishing, 2015.
12. Malvey, D.; Slovensky, D.J. mHealth: Transforming Healthcare, 1st ed.; Springer US, 2014.
13. Lang, M. Heart Rate Monitoring Apps: Information for Engineers and Researchers About the New European Medical Devices Regulation 2017/745. JMIR Biomed Eng **2017**, 2, e2.
14. Quinn, P. The EU commission's risky choice for a non-risk based strategy on assessment of medical devices. Computer Law & Security Review **2017**, 33, 361–370.
15. Sperlich, B.; Holmberg, H.C. Wearable, yes, but able...?: it is time for evidence-based marketing claims! Br J Sports Med Published Online First: 16 December 2016, doi: 10.1136/bjsports-2016-097295.
16. Lang, M. Beyond Fitbit: A Critical Appraisal of Optical Heart Rate Monitoring Wearables and Apps, their Current Limitations and Legal Implications. Alb. L. J. Sci. & Tech. **2018**, 28. forthcoming.
17. Tripathi, O.N.; Ravens, U.; Sanguinetti, M.C., Eds. Heart Rate and Rhythm: Molecular Basis, Pharmacological Modulation and Clinical Implications, 1st ed.; Springer Berlin Heidelberg, 2011.
18. Jangra, M.; Dhull, S.K.; Singh, K.K., Recent trends in arrhythmia beat detection: A review. In Communication and Computing Systems; Prasad, B.; Singh, K.K.; Ruhil, N.; Singh, K.; O'Kennedy, R., Eds.; Taylor & Francis Group, CRC Press, 2016; pp. 177–183.
19. Veeravalli, B.; Deepu, C.J.; Ngo, D., Real-Time, Personalized Anomaly Detection in Streaming Data for Wearable Healthcare Devices. In Handbook of Large-Scale Distributed Computing in Smart Healthcare; Khan, S.U.; Zomaya, A.Y.; Abbas, A., Eds.; Springer International Publishing, 2017; pp. 403–426.
20. Elgendi, M.; Eskofier, B.; Dokos, S.; Abbott, D. Revisiting QRS Detection Methodologies for Portable, Wearable, Battery-Operated, and Wireless ECG Systems. PLOS ONE **2014**, 9, 1–18.
21. Lemkaddem, A.; Proença, M.; Delgado-Gonzalo, R.; Renevey, P.; Oei, I.; Montano, G.; Martinez-Heras, J.A.; Donati, A.; Bertschi, M.; Lemay, M. An autonomous medical monitoring system: Validation on arrhythmia detection. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017, pp. 4553–4556.
22. Ye, C.; Coimbra, M.T.; Kumar, B.V.K.V. Arrhythmia detection and classification using morphological and dynamic features of ECG signals. 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, 2010, pp. 1918–1921.
23. Chang, R.C.H.; Lin, C.H.; Wei, M.F.; Lin, K.H.; Chen, S.R. High-precision real-time premature ventricular contraction (PVC) detection system based on wavelet transform. Journal of Signal Processing Systems **2014**, 77, 289–296.
24. Yaghouby, F.; Ayatollahi, A.; Bahramali, R.; Yaghouby, M.; Alavi, A. Towards automatic detection of atrial fibrillation: a hybrid computational approach. Comput. Biol. Med. **2010**, 40, 919–930.
25. Acharya, U.; Bhat, B.; Iyengar, S.; Rao, A.; Dua, S. Classification of heart rate data using artificial neural network and fuzzy equivalence relation. Pattern Recognit. **2003**, 36, 61–68.
26. Kumar, S.; Bansal, A.; Tiwari, V.; Nayak, M.; Narayanan, R. Remote health monitoring system for detecting cardiac disorders. Proceedings of 2014 IEEE International Conference on Bioinformatics and Biomedicine, 2014, pp. 30–34.
27. Gradl, S.; Kugler, P.; Lohmüller, C.; Eskofier, B. Real-time ECG monitoring and arrhythmia detection using Android-based mobile devices. 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2012, pp. 2452–2455.

28. Luz, E.J.S.; Schwartz, W.R.; Cámara-Chávez, G.; Menotti, D. ECG-based heartbeat classification for arrhythmia detection: A survey. Computer Methods and Programs in Biomedicine **2016**, 127, 144–164.
29. Oresko, J.J.; Jin, Z.; Cheng, J.; Huang, S.; Sun, Y.; Duschl, H.; Cheng, A.C. A Wearable Smartphone-Based Platform for Real-Time Cardiovascular Disease Detection Via Electrocardiogram Processing. IEEE Transactions on Information Technology in Biomedicine **2010**, 14, 734–740.
30. Amiri, A.M.; Abhinav.; Mankodiya, K. m-QRS: An efficient QRS detection algorithm for mobile health applications. 2015 17th International Conference on E-health Networking, Application Services (HealthCom), 2015, pp. 673–676.
31. Kim, Y.J.; Heo, J.; Park, K.S.; Kim, S. Proposition of novel classification approach and features for improved real-time arrhythmia monitoring. Computers in biology and medicine **2016**, 75, 190–202.
32. Tsipouras, M.G.; Fotiadis, D.I. Automatic arrhythmia detection based on time and time–frequency analysis of heart rate variability. Computer Methods and Programs in Biomedicine **2004**, 74, 95–108.
33. de Chazal, P.; O'Dwyer, M.; Reilly, R.B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Transactions on Biomedical Engineering **2004**, 51, 1196–1206.
34. Pan, J.; Tompkins, W.J. A Real-Time QRS Detection Algorithm. IEEE Transactions on Biomedical Engineering **1985**, BME-32, 230–236.
35. Moskvina, V. Application of the Singular Spectrum Analysis for change-point detection in time serie. PhD thesis, UK, 2001. Cardiff University.
36. Moskvina, V.; Zhigljavsky, A. An Algorithm Based on Singular Spectrum Analysis for Change-Point Detection. Communications in Statistics **2003**, 32, 319–352.
37. Golyandina, N.; Nekrutkin, V.; Zhigljavsky, A. Analysis of Time Series Structure: SSA and related techniques, 1st ed.; Chapman & Hall/CRC Monographs on Statistics & Applied Probability (Book 90), Chapman and Hall/CRC, 2001.
38. Dong, Q.; Yang, Z.; Chen, Y.; Li, X.; Zeng, K. Anomaly Detection in Cognitive Radio Networks Exploiting Singular Spectrum Analysis. International Conference on Mathematical Methods, Models, and Architectures for Computer Network Security. Springer, 2017, pp. 247–259.
39. Yang, Z.; Zhou, N.; Polunchenko, A.; Chen, Y. Singular Spectrum Analysis Based Quick Online Detection of Disturbance Start Time in Power Grid. 2015 IEEE Global Communications Conference (GLOBECOM), 2015, pp. 1–6.
40. Zhang, S.; Liu, Y.; Pei, D.; Chen, Y.; Qu, X.; Tao, S.; Zang, Z.; Jing, X.; Feng, M. FUNNEL: Assessing Software Changes in Web-based Services. IEEE Transactions on Services Computing **2016**, PP.
41. Georgescu, V.; Delureanu, S.M. Fuzzy-valued and complex-valued time series analysis using multivariate and complex extensions to singular spectrum analysis. 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2015, pp. 1–8.
42. Jarchi, D.; Lo, B.; Wong, C.; Ieong, E.; Nathwani, D.; Yang, G.Z. Gait Analysis From a Single Ear-Worn Sensor: Reliability and Clinical Evaluation for Orthopaedic Patients. IEEE Transactions on Neural Systems and Rehabilitation Engineering **2016**, 24, 882–892.
43. Uus, A.; Liatsis, P. Singular Spectrum Analysis for detection of abnormalities in periodic biosignals. 2011 18th International Conference on Systems, Signals and Image Processing, 2011, pp. 1–4.
44. Sanei, S.; Hassani, H. Singular Spectrum Analysis of Biomedical Signals, 1st ed.; CRC Press, 2016.
45. Clifford, G.; Silva, I.; Moody, B.; Li, Q.; Kella, D.; Shahin, A.; Kooistra, T.; Perry, D.; Mark, R. The PhysioNet/Computing in Cardiology Challenge 2015: Reducing False Arrhythmia Alarms in the ICU. Computing in cardiology **2015**, 42, 273–276.
46. Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. Physiobank, physiotoolkit, and physionet. Circulation **2000**, 101, e215–e220.
47. Golyandina, N.; Zhigljavsky, A. Singular Spectrum Analysis for Time Series, 1st ed.; Springer Briefs in Statistics, Springer, 2013.
48. Elsner, J.B.; Tsonis, A.A. Singular Spectrum Analysis: A New Tool in Time Series Analysis; Language of Science, Springer, 1996.
49. Idé, T.; Inoue, K. Knowledge discovery from heterogeneous dynamic systems using change-point correlations. Proceedings of the 2005 SIAM International Conference on Data Mining. SIAM, 2005, pp. 571–575.
50. Mohammad, Y.; Nishida, T. Robust Singular Spectrum Transform. Next-Generation Applied Intelligence: 22nd International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE

2009, Tainan, Taiwan, June 24-27, 2009. Proceedings; Chien, B.C.; Hong, T.P.; Chen, S.M.; Ali, M., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; pp. 123–132.

51. Idé, T.; Tsuda, K. Change-Point Detection using Krylov Subspace Learning. Proceedings of the 2007 SIAM International Conference on Data Mining, 2007, pp. 515–520, [http://epubs.siam.org/doi/pdf/10.1137/1.9781611972771.54].
52. McDonald, D.R. A Cusum Procedure Based on Sequential Ranks; Laboratory for Research in Statistics and Probability, Carleton University, 1985.
53. Pflugradt, M.; Geissdoerfer, K.; Goernig, M.; Orglmeister, R. A Fast Multimodal Ectopic Beat Detection Method Applied for Blood Pressure Estimation Based on Pulse Wave Velocity Measurements in Wearable Sensors. Sensors **2017**, 17.
54. Salvi, P. Pulse Waves: How Vascular Hemodynamics Affects Blood Pressure, 1st ed.; Springer, 2012.
55. Page, E.S. Continuous Inspection Schemes. Biometrika **1954**, 41, pp. 100–115.
56. Lucas, J.M. The Design and Use of Cumulative Sum Control Schemes. Technometrics **1976**, pp. 51–59.
57. Chatterjee, S.; Qiu, P. Distribution-free cumulative sum control charts using bootstrap-based control limits. Ann. Appl. Stat **2009**, 3, 349–369.
58. Reynolds, M.R. Approximations to the average run length in cumulative sum control charts. Technometrics **1975**, pp. 65–71.