Review of Motion Artifacts Removing Techniques for Wireless Electrocardiograms

Waltenegus Dargie and Jannis Lilienthal
Faculty of Computer Science, Technische Universität Dresden, 01062 Dresden, Germany
Email: {waltenegus.dargie, jannis.lilienthal}@tu-dresden.de

***Abstract*—A certain class of wireless devices such as electrocardiogram (ECG), electromyogram (EMG) and electroencephalogram (EEG), are very useful for telemedicine because they enable the free movement of patients while vital biophysical measurements are taken from them. However, these devices are very sensitive to motion artifacts – electric potentials generated due to the undesirable movement of electrodes on the surface of the skin or the change in the skin impedance. In this paper we examine the scope and usefulness of different types of model-based signal processing and dimensionality reduction techniques to model and reason about motion artifacts. While the techniques we review are applicable for a wide range of signals, we limit our analysis, nevertheless, to wireless electrocardiograms, so that we can base our investigation on experimental data.**

***Index Terms*—Adaptive filters, Electrocardiogram, motion artifacts, Independent Component Analysis, Singular Value Decomposition , Tensor Decomposition**

I. Introduction

Wireless medical devices such as wireless electrocardiogram (ECG), wireless electromygram (EMG) and wireless electrocardiogram (EEG) measure action potentials on the surface of the skin using different types of electrodes. Action potentials are electrical voltages generated by the excitement of tissue, nerve and muscle cells. The ECG measures cardiac action potentials; the EMG, action potentials generated by the muscles; and the EEG, action potentials generated by the neurons of the cerebral cortex [1], [2]. The devices are useful for diagnosing and monitoring a plethora of diseases, but their application has so far been limited to clinical settings, mainly due to their sensitivity to motion.

If we consider just two ECG electrodes at the surface of the skin to the right and left of the sternum, they essentially measure a potential difference between two points. As the distance between the electrodes varies, so does the potential difference (partly, because the impedance between the electrodes is distance dependent). Hence, the undesired movement (displacement) of one of the electrodes with respect to the other introduces a spurious change in the voltage measurement which can be mistaken for a change in the underlying cardiac action potentials. Because this change arises due to the movement of the electrode(s), it is referred to as a motion artifact [1].

In order to remove motion artifacts, a reference signal or statistics pertaining to either the useful signal or the motion artifacts should be available. However, a reference model of the useful signal cannot be established, as it depends on the underlying cardiac, muscular, or brain conditions, which are unknown. Similarly, the motion affecting the electrodes is a result of complex and time-varying physical movements, so that no readily available model can be employed to estimate their statistics. In the literature, attempts have been made to establish a correlation between the physical movements affecting the electrodes and the useful signal by using accelerometers, tilt, or contact sensors [3], [4], [5]. Some have employed infrared sensors to indirectly establish the movement of electrodes from the way the skin texture changes when it contracts and relaxes in response to the movement of the body [6]. Some have embedded additional circuits into the EEG/ECG/EMG platforms in order to measure the change in the electrical characteristics of the interface between the electrodes and the skin [7], [8], [9]. Regardless of the specific sensors employed, the idea is to estimate the motion artifacts in terms of the motion affecting the electrodes:

$$\hat{\mathbf{n}} = g\left(\mathbf{a}\right) \tag{1}$$

where $\hat{\mathbf{n}}$ is the estimation of the motion artifacts $\mathbf{n}$ in terms of the physical motion represented by $\mathbf{a}$ and $g(\bullet)$ is some mapping function. All the parameters are regarded as random variables, as they do not accurately represent the underlying physical conditions. Furthermore, besides the external (undesirable) motion affecting the electrodes, the internal activities themselves induce their own motion on the electrodes. In the case of ECG, cardiac action potentials cause the myocardia to contract and relax thereby making the heart pump blood; in EMG, the action potentials are results of the actual contractions and relaxations of muscles. In both cases, the action potentials are associated with actual underlying movements which can be picked by the motion sensors. The purpose of this paper is to closely examine the scope and usefulness of some of the proposed techniques to remove motion artifacts and to examine the validity of the underlying assumptions. The rest of this paper is organized as follows: In Section II we give account to how we obtained the data we use for our investigation. In Sections III and IV we review two probabilistic estimation techniques, namely, Independent Component Analysis and Adaptive Filters. In Sections V and VI, we review two dimensionality reduction techniques, namely, Singular Value Decomposition and Tensor Decomposition. Finally, in Section VII we give concluding remarks.

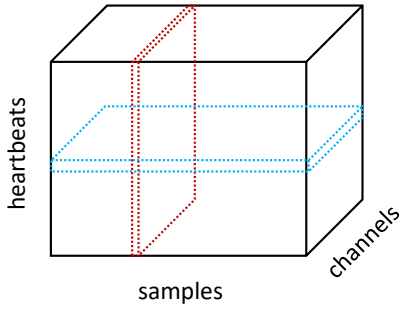

Fig. 1. A three-dimensional array consisting of the raw data from a 5-lead wireless ECG and 2 inertial measurement units (3D accelerometer and 3D gyroscope). The array belongs to a single movement lasting approximately 10 s.

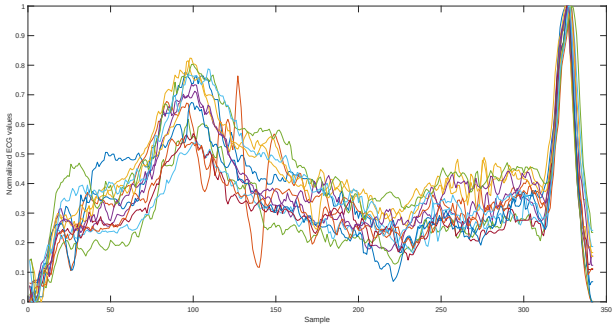

Fig. 2. Thirteen successive ECG segments measured between the left arm and the right arm (LA-RA) of a subject while jumping. The measurement duration was approximately 10 s.

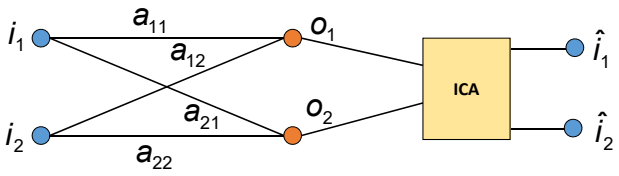

Fig. 3. The model parameters in Independent Component Analysis (ICA)

II. Background

For our evaluation, we rely on actual experiments we conducted using the Shimmer-3 Platform consisting of a 5-lead wireless electrocardiogram, a 3D accelerometer, and a 3D gyroscope [10], among others. The platform allows the synchronous sampling of all on-board sensors. We set the sampling rate at 512 samples per second and saved the data locally. The measurements were taken from 11 university employees doing six simple exercises (Bending forward, Standing up, Climbing stairs, Walking, Running, and Jumping) and all the experiments were conducted with the approval of the TU Dresden's Ethics Committee (Reference Number: EK 271072017).

Under normal circumstances, the samples of the ECG are correlated in two complementary ways. Firstly, the heart beats fairly regularly. Secondly, within a single heartbeat, there is a conditional dependency between the occurrence of the **P** wave, the **QRS** wave complex, the **T** wave and the segments in between. Even in the absence of any external reference signal, these dependencies can be exploited to estimate the presence and extent of motion artifacts, provided that there are no underlying cardiac conditions. Nevertheless, there is some irregularity in the way the heart beats, bringing uncertainty in the estimation assignment. This uncertainty increases when motion artifacts and underlying cardiac conditions are mixed. Some of the proposed approaches attempt to establish a correlation between the motion artifacts and the reference signals in the absence of any underlying cardiac conditions – i.e., by monitoring healthy subjects in motion. One way to achieve this is to segment the row ECG measurements into heartbeats and set up a three-dimensional array (heartbeats vs. samples vs. channels), as shown in Fig. 1.

Fig. 2 shows thirteen successive ECG segments lasting approximately 10 s. The measurements were taken while one of our subjects was jumping. As can be seen, the **QRS** wave complex (the waveform at the right end) are the least affected by motion artifacts, mainly due to their distinct and strong nature whereas all the others are affected to some extent.

The most widely employed techniques for removing motion artifacts are Independent Component Analysis (ICA) [11], [12], [13], [14] and Adaptive Filters (AF) [15], [16], [17], Singular Value Decomposition [18], [19], [20], Principal Component Analysis (PCA) [21], [22], and tensor decomposition [23], [24], [25], [26]. In the subsequent sections we review these approaches closely. Nevertheless we do not review Principal Component Analysis. This approach is essentially similar to SVD and repeated experiments yielded identical results.

III. Independent Component Analysis

Arising originally in the field of audio signal processing, the Independent Component Analysis (ICA) aims to separate two or more independent source signals which are supposed to have been linearly mixed. There are three basic underlying assumptions for ICA to yield reliable results:

- The input sources are, statistically speaking, independent.
- There are at least as many output signals as source signals.
- In case of two source signals, at least one of them should have a non-Gaussian (non-normal) density function.

Fig. 3 displays the basic estimation model for ICA. Two independent signal sources (i_1 and i_2), for our case, the true ECG signal and the motion artifacts, are mixed with one another. These signals are picked by two independent detectors (sensors) producing o_1 (the noisy EEG/EMG/ECG signal) and o_2 (the signal picked up by the motion sensor). The source and the output signals are all regarded as random variables having their own probability density functions, $f(i_1), f(i_2), f(o_1)$ and $f(o_2)$. Furthermore, it is assumed that the output signals are mixed linearly, so that:

$$\mathbf{o}_1 = a_{11}i_1 + a_{12}i_2 \tag{2}$$

$$\mathbf{o}_2 = a_{21}i_1 + a_{22}i_2 \tag{3}$$

or,

$$\begin{bmatrix} \mathbf{o}_1 \\ \mathbf{o}_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \tag{4}$$

or, in short:

$$\mathbf{o} = \mathbf{A}\mathbf{i} \tag{5}$$

where $\mathbf{i}$ is the input or source vector containing i_1 and i_2 and $\mathbf{A}$ is the mixing matrix. From which, we have:

$$\mathbf{i} = A^{-1}\mathbf{o} = \mathbf{W}\mathbf{o} \tag{6}$$

where $\mathbf{A}^{-1} = \mathbf{W}$ is the inverse matrix which is also called unmixing matrix. The corresponding density functions are related with one another as follows:

$$f(o) = f(i)\,|\mathbf{W}^*| \tag{7}$$

where $|\mathbf{W}^*| = \partial\mathbf{i}/\partial\mathbf{o}$ is the Jacobian matrix of $\mathbf{o}$ with respect to $\mathbf{i}$.

The aim of ICA is to reproduce i_1 and i_2 by first estimating $\mathbf{W}$. The estimation steps/conditions are as follows:

1) Because one has already measured o_1 and o_2, their statistics are readily available. Furthermore, since each of these random variables arises from the summation of two random variables (i_1 and i_2) , their density functions are approximated to be Gaussian, in accordance with the Central Limit Theorem [27] (Ch. 7; refer also to [28] for more information).
2) Even though i_1 and i_2 are unknown random variables, their density functions, nevertheless, are supposed to be known (for one of them the condition of non-Gaussian distribution must hold). If, however, both of them are Gaussian, their summation, too, will be Gaussian and separating the two signals from one another without any additional knowledge will not be possible.
3) But if one of them is non-Gaussian, then, the ICA first attempts to estimate the inverse matrix which maximises the output signals' statistics given the statistics of the input signals. Often, the Maximum Likelihood Estimation is employed for this task [29]. Hence,

$$f(o|\mathbf{W}) = f(\mathbf{W}\mathbf{i})\,|\mathbf{W}| \tag{8}$$

Equation (8) is expected to be maximum when $\mathbf{W} = \mathbf{W}^*$ in Equation (7). Consequently, Equation (8) is often employed to evaluate the eligibility of any putative unmixing matrix $\mathbf{W}$ which maximizes $f(o_1|\mathbf{W})$.
4) Once $\mathbf{W}$ is determined, then the values of i_1 and i_2 can be estimated sample by sample.

In plain terms, consider a single value observed at one of the outputs, say at $\mathbf{o}_1$, $\mathbf{o}_1 = v_1^1$. Since we have said that the output is the superposition of two independent inputs, we have an infinite combinations of values that might have resulted in v_1^1:

$$\begin{aligned} v_1^1 &= i_1^a + i_2^\alpha \\ v_1^1 &= i_1^b + i_2^\beta \\ v_1^1 &= i_1^c + i_2^\gamma \\ &\vdots \end{aligned}$$

However, since we already have assumed that statistics pertaining to the input signals are already available, we know with what probabilities $i_1^a, i_1^b, i_1^c, i_2^\alpha, i_2^\beta, i_2^\gamma$, etc. occur. Similarly, from the output statistics, we know the probability of v_1^1. Moreover, from our independence assumption, we know that $p(v_1) = p(i_1^a)p(i_2^\alpha) + p(i_1^b)p(i_2^\beta) + \cdots$. However, at any given instance, only one of the input combinations is possible. Therefore, the maximum likelihood estimate attempts to pick up one of those combinations which are the most likely or which maximize the probability of v_1^1.

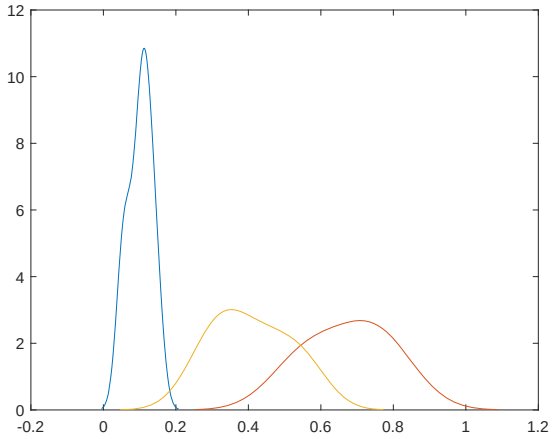

Fig. 4. The distributions of three sample points (50, 100, and 150) in an ECG taken between the left arm and the right arm (LA-RA) while a subject was jumping.

Evaluation

If we take each sample point of a heartbeat as a random variable, the distribution of these random variables can be established by considering the values of the successive heartbeats for the same sample points. Fig. 4 shows the distributions of the sample points 50, 100, and 150 for the heartbeats shown in Fig. 2. A narrower curve implies that the sample points of the successive segments are consistent with one another and, hence, less affected by motion artifacts. The corresponding measurement taken synchronously from one of the accelerometer axes is shown in Fig. 5

In ICA, it is assumed that since the output random variables are mixtures of two independent inputs (for our case), their distributions should be Gaussian. If we take any one of the random variables in Fig. 4 as o_1 and the corresponding random variable in Fig. 5 as o_2 to set up Equation (5), then the assumption can be taken as plausible. It is apparent that both output random variables are normally distributed.

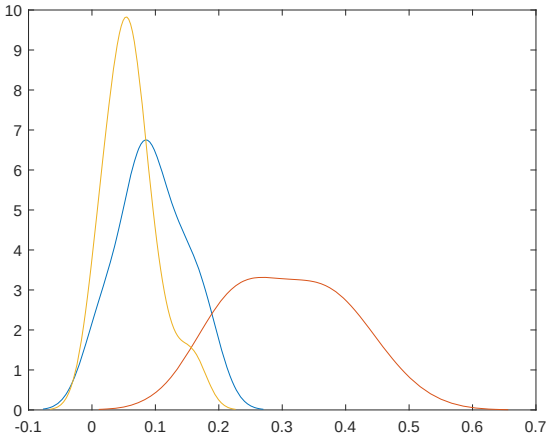

Fig. 5. The distributions of three sample points taken from one of the accelerometers (the y-axis).

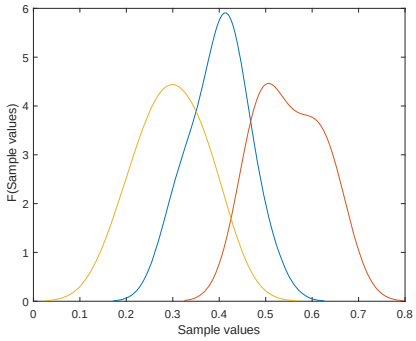

Fig. 6. The distributions of three ECG sample points (50, 100, and 150) taken under an ideal condition.

Fig. 6 displays the distributions of the same ECG samples taken in an ideal condition (i.e., with minimal motion artifacts present). These random variables can be taken as the inputs to Equation (5). The randomness in the sample values can be attributed to the irregularities of the heartbeats even under normal circumstances, in which case the distribution of the artifacts-free ECG at any point can be taken as a normally distributed random variable.

We have now seen that at least three of the parameters in the artifacts removal assignment are potentially normally distributed. What we do not know is the distribution of the motion artifacts, one of the input parameters. In the absence of this knowledge, it is difficult to justify the use of ICA.

IV. Adaptive Filter

The implementation of an *Adaptive Filter* consists of three tasks: Prediction, measurement, and combination. These tasks may refer to the estimation of either the ECG or the motion artifacts. In the first case, the motion artifacts are considered only indirectly whereas in the second, they are the subject of estimation, in which case, another step is required to estimate the clean ECG as follows:

$$\hat{\mathbf{m}} = \mathbf{r} - \hat{\mathbf{n}} \tag{9}$$

where $\hat{\mathbf{m}}$ refers to the best estimation of the clean ECG; $\mathbf{r}$, to the noisy ECG; and $\hat{\mathbf{n}}$, to the estimated motion artifacts. Whether the ECG or the motion artifacts are estimated, in the Adaptive Filter, the difference between the desired signal and its estimate is minimized in a mean square error sense [30].

Suppose we wish to estimate the motion artifacts in a noisy ECG using the outputs of a 3D accelerometer (this refers to the measurement task):

$$\hat{\mathbf{n}} = \alpha_x \mathbf{a}_x + \alpha_y \mathbf{a}_y + \alpha_z \mathbf{a}_z \tag{10}$$

Hence, the mean square error is given as:

$$E\left\{\mathbf{e}^2\right\} = E\left\{\left(\mathbf{n} - \left[\alpha_x \mathbf{a}_x + \alpha_y \mathbf{a}_y + \alpha_z \mathbf{a}_z\right]\right)^2\right\} \tag{11}$$

The error in Equation (11) can be minimized by finding the optimal α_x, α_y, and α_z. This can be done by differentiating the equation with respect to α_x, α_y, and α_z, and setting the result to zero, which yields:

$$\begin{array}{rcl} E\left\{\mathbf{n}\mathbf{a}_x\right\} & = & \alpha_x E\left\{\mathbf{a}_x^2\right\} + \alpha_y E\left\{\mathbf{a}_x\mathbf{a}_y\right\} + \alpha_z E\left\{\mathbf{a}_x\mathbf{a}_z\right\} \\ E\left\{\mathbf{n}\mathbf{a}_y\right\} & = & \alpha_x E\left\{\mathbf{a}_x\mathbf{a}_y\right\} + \alpha_y E\left\{\mathbf{a}_y^2\right\} + \alpha_z E\left\{\mathbf{a}_y\mathbf{a}_z\right\} \\ E\left\{\mathbf{n}\mathbf{a}_z\right\} & = & \alpha_x E\left\{\mathbf{a}_x\mathbf{a}_z\right\} + \alpha_y E\left\{\mathbf{a}_y\mathbf{a}_z\right\} + \alpha_z E\left\{\mathbf{a}_z^2\right\} \end{array} \tag{12}$$

Equation (12) can be expressed in a matrix form:

$$\mathbf{R}_{na} = \alpha \mathbf{R}_{aa} \tag{13}$$

where $\mathbf{R}_{na}$ refers to the left term in Equation (12) expressing the correlation between the motion artifacts and the linear acceleration picked by the axes of the accelerometer and $\mathbf{R}_{aa}$ is the correlation between the different axes of the accelerometer. From which we have:

$$\alpha = \mathbf{R}_{na}\mathbf{R}_{aa}^{-1} \tag{14}$$

Notice that $\mathbf{R}_{aa}$ contains only measurable quantities which can readily be available after an experiment. Equations (10) – (14) essentially summarize the assumptions one makes when applying an Adaptive Filter:

1) To begin with, from $\mathbf{R}_{na}$, it is clear that the measurements of the 3D accelerometer are assumed to be correlated with the motion artifacts included in the noisy ECG signal. Often a reference setting (such as controlled movements) is used to determine this correlation. In general, however, there is no direct mechanism to ascertain this correlation and this is one of the limitations of using an Adaptive Filter.
2) The weight assigned to a given acceleration component is proportional to its perceived correlation with the motion artifacts ($\mathbf{R}_{na}$) and inversely proportional to the relative uncertainty it introduces (i.e., its variance). This fact, which is encoded in $\mathbf{R}_{aa}^{-1}$, is intuitive: The higher our uncertainty in a random variable, the lesser should be our confidence in it[1].

[1]The measure of uncertainty in a random variable is expressed by its variance. The more dissimilar the samples of a random variable are, the bigger its variance, and, therefore, the bigger is the uncertainty in its outcomes.

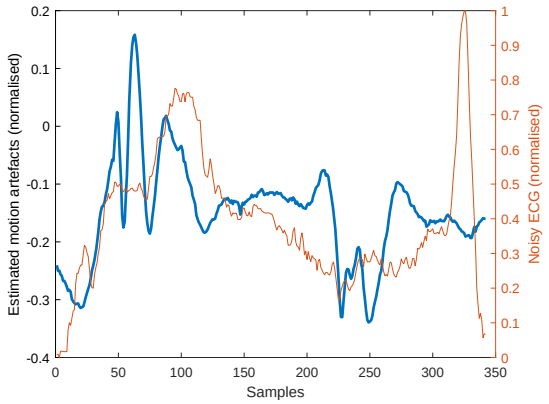

Fig. 7. The estimation of motion artifacts embedded in a noisy ECG using an Adaptive Filter. The reference signals are obtained from a 3D accelerometer.

Evaluation

Fig. 7 compares a noisy ECG (coral) and the estimation of the motion artifacts embedded using an *Adaptive Filter*. Measurements from a 3D accelerometer are used as reference for the estimation. The correlation between the three axes and the ECG is used to compute $\mathbf{R}_{na}$. From the plots it is possible to relate the distortion in the ECG with the physical motion that is picked by the 3D accelerometer.

V. Singular Value Decomposition

Suppose we have m sources jointly "observing" certain phenomena of interest, each imperfect in its own way. If we sample the sources synchronously (i.e., time being a common variable), then the samples taken from these sources can make up a matrix, the rows making up the samples and the columns making up the sources (the random variables). If the matrix is a square matrix, then it is possible to investigate its eigenvectors and eigenvalues to determine the unique features the samples describe. An interesting quality of the eigenvectors is that they are orthonormal, which means, statistically speaking, they are independent. For our case, the following observations are useful to notice:

- The ECG samples are correlated with each other. This correlation will be taken into account during decomposition.
- The motion artifacts are independent of the underlying ECG, as they are results of the motion the electrodes experience on the surface of the skin. In other words, the motion artifacts and the ECG will be regarded as independent features during a decomposition.
- The measurement sources are independent of one another and, therefore, the measurement error introduced by each source will be considered as an independent feature.

Any rectangular matrix $\mathbf{A}$ of size $n \times m$ can be decomposed into three matrices: $\mathbf{U}$ $(n \times r)$, Σ $(r \times r)$, and $\mathbf{V}$ (m $\times r$) without any loss of information:

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^{\mathbf{T}} \tag{15}$$

$\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices[2] and Σ is a non-negative diagonal matrix whose entries $\sigma_{ii}, 1 \leq i \leq r$ are sorted in decreasing order. The diagonal entries of Σ are called the singular values of $\mathbf{A}$ and encode its distinct features and their relative significance. Each column of $\mathbf{U}$ encodes a single feature which is independent of all the other features and ranks the rows of $\mathbf{A}$ in terms of this feature. Similarly, each column of $\mathbf{V}$ encodes a single feature and ranks the columns of $\mathbf{A}$ in terms of this feature. The significance of the features, as we already remarked, is encoded in Σ.

The advantage of SVD is that few assumptions need to be made as regards the correlation between the measurement sets (as random variables) and their probability distributions. SVD automatically separates the hidden features into independent, non-overlapping feature spaces. The disadvantage, on the other hand, is that there are no formal steps which guarantee the objective interpretation of $\mathbf{U}$ and $\mathbf{V}$. Often a case by case analysis is required to make sense of these matrices.

Evaluation

Fig. 8 displays the plots of the $\mathbf{U}$ and the $\mathbf{V}$ matrices associated with the SVD of a matrix containing a noisy ECG and measurements from a 3D accelerometer and a 3D gyroscope for a single heartbeat. The measurements were taken while a subject was jumping. The original matrix had 342 rows and 9 columns, the rows representing the samples and the columns, the sources. The first 3 columns refer to the $x, y,$ and z axes of the accelerometer; the columns $4, 5$ and 6 refer to the three ECG channels (LL-LA, LL-RA, and LA-RA), and the last three columns refer to the $x, y,$ and z axes of the gyroscope.

The analysis of Fig. 8 requires domain knowledge. The plots on the left side display the columns of the $\mathbf{V}$ matrix encoding the contribution of each source in describing a specific feature. Likewise, the plots on the right side display the columns of the $\mathbf{U}$ matrix encoding how each feature manifests itself at a specific sample point. It should be recalled that the accelerometer and the gyroscope could pick up the external motion of the body and the internal motion due to the pumping of the heart. Where sources scored comparably in the $\mathbf{V}$ plot, we can assume that they agree in their observation. Where the scores show significant variations, then there was no agreement between them suggesting that the phenomenon was unique to a specific source. This phenomenon is very likely an uncorrelated noise.

Thus, two interesting phenomena are picked up by all the sources, the first occurred approximately at the 51st sample point and the second at the 224th. These two aspects were also estimated by the Adaptive Filter in Fig. 7 where we attributed them to motion artifacts. Since no ECG wave forms resemble them, indeed these two phenomena must be motion artifacts. Exception to this observation are the fifth and sixth right-side plots from the top. In order to analyze these plots, we should

[2]The matrix $\mathbf{U}$ is said to be orthonormal if and only if $\mathbf{U}^{\intercal}\mathbf{U} = \mathbf{I}$ or, alternatively, $\mathbf{U}^{-1} = \mathbf{U}^{\intercal}$.

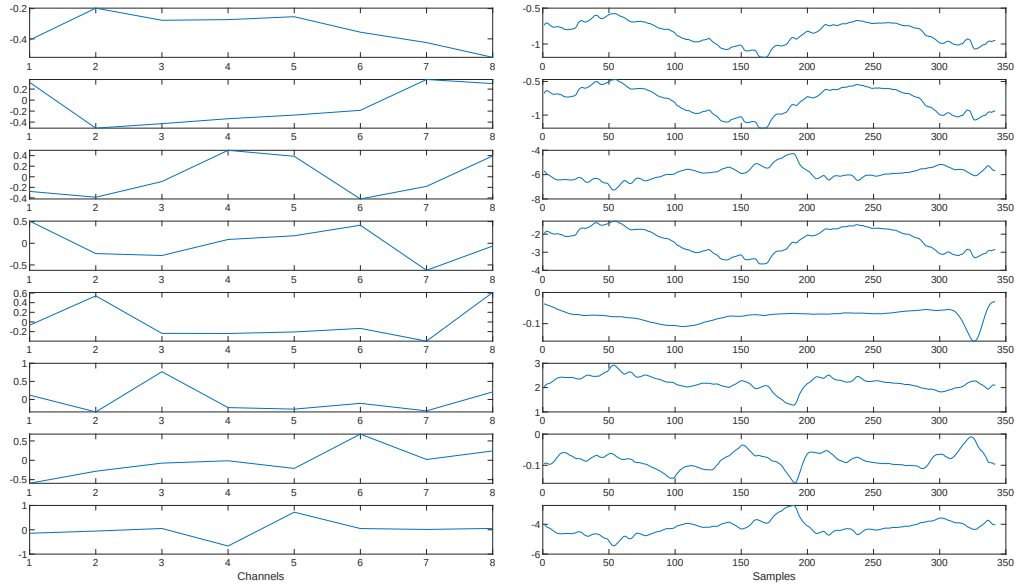

Fig. 8. Using SVD to decompose a noisy ECG and the output of six inertial measurement units. The y-axis in all the graphs refers to the SVD scores of the channels/samples in the corresponding features.

first refer to the corresponding left-side plots. In the fifth plot, the channels which scored comparable results are channel 4, 5, and 6, all of which are ECG channels. Hence, we can presume this feature to belong to the ECG. Indeed, a close examination reveals a swelling around the sample point 100 which clearly belongs to the **T** wave and the unmistakable **QRS** wave at sample point 325. By contrast, the sources in the sixth plot (left) have variable scores. Therefore, the right side plot must belong to an uncorrelated noise.

VI. Tensor Decomposition

One of the limitations of SVD is that it works only for a matrix (two dimensions). Tensor decomposition can increase the degree of freedom by at least one dimension. Every entry of a matrix x_{ij} is an intersection between the i-th row and the j-th column. An n-way tensor (or an n-dimensional array) is one whose entries are referred by n indices. A matrix is sometimes referred to as a two-way tensor. The measurement set displayed in Fig. 1 is referred to as a three-way tensor (heartbeats vs. channels vs. samples).

In the same way a matrix can be decomposed into basic constituting elements, a tensor, too, can be decomposed into basic constituting elements. However, unlike decomposing a matrix, decomposing a tensor is not straightforward. To start with, an assumption has to be made about the number of features that are hidden in the original tensor, whereas this is done automatically with SVD. Secondly, different decomposition strategies employ different statistical estimation techniques and decomposition structures, which may result in different outcomes.

A closer look into the tensor in Fig. 1 reveals that it provides three orthogonal views which can serve different purposes. For example, the front view provides a matrix describing the relationship between the heartbeats and the samples for the k-th source (channel) – i.e, (heartbeats vs. samples)$_k$. This view is called the front slice. Likewise, the top view provides a matrix describing the relationship between the samples and the channels for the i-th heartbeat – i.e., (channels vs. samples)$_i$. This is called the horizontal slice. Finally, the side view provides a matrix describing the relationship between the heartbeats and the channels for the j-th sample– i.e., (heartbeats vs. channels)$_j$. This is called the lateral slice. It is this flexibility, among others, which makes a tensor desirable. The chief task of a tensor decomposition is to explore multidimensional correlations in terms of which the original tensor can be explained. Compared to the size of the tensor, the basic features are typically small if the samples are correlated with each other. For our case, we assume that the outputs of the motion sensors are to some extent correlated with the ECG as well as the motion artifacts embedded in it. So, the decomposition should uncover these correlations.

A tensor decomposition begins by unfolding (flattening) the tensor into a matrix. The unfolding takes place in different modes, but whichever way is chosen, the entries along each dimension form a column vector. The mode-1 unfolding of the above tensor takes each column vector as they are and put them together side-by-side. The mode-2 unfolding takes each raw entry of the matrices and places them as column vectors in a single matrix. Likewise, the mode-3 unfolding takes the n-th entry of each slice (along the time dimension) and puts them together as column matrix (numerical examples can be found in [23], [31]).

There are different tensor decomposition strategies. Which of them is the most suitable depends on the application. The easiest to interpret and the most widely employed is the Canonical Decomposition/Parameter Factorization (referred in

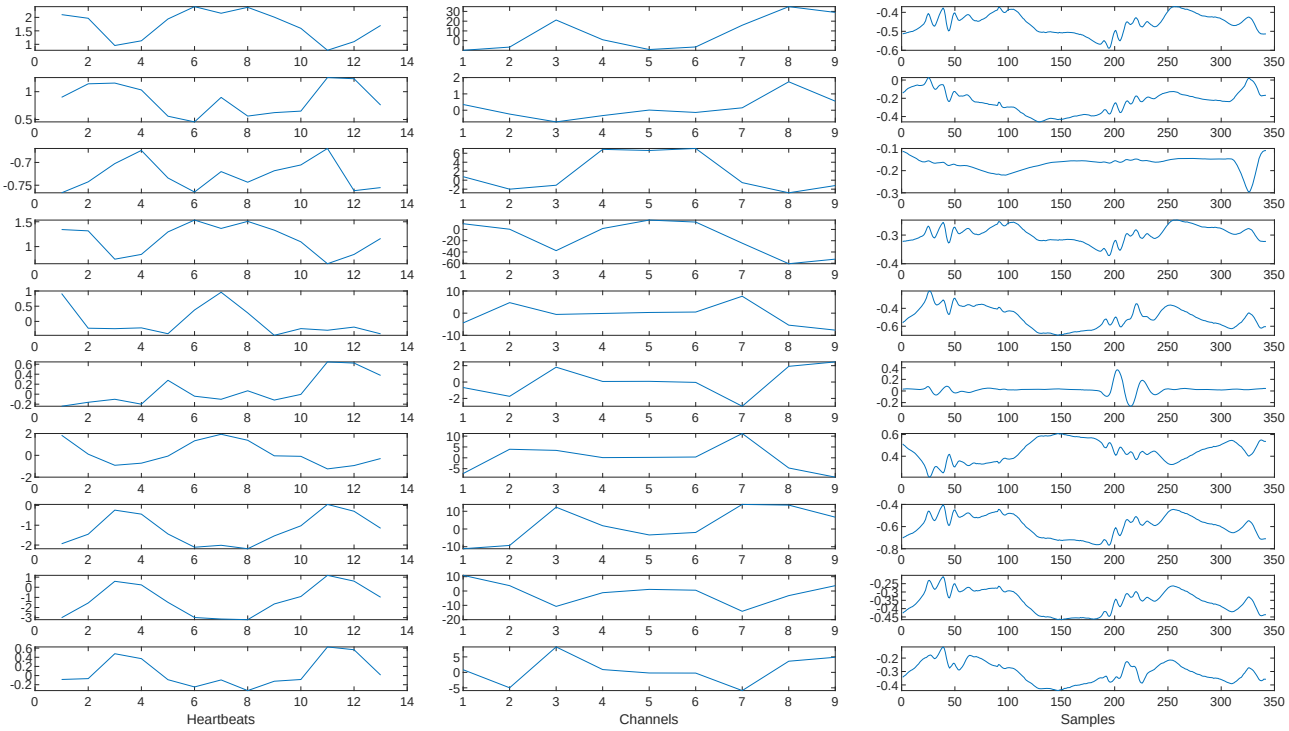

Fig. 9. The results of decomposing a tensor containing measurements of three ECG leads (channels), three dimensional acceleration, and three dimensional angular velocity. The CANDECOMP/PARAFAC decomposition technique is employed. The y-axis in all the graphs refers to the scores of the heartbeats/channels/samples in describing a corresponding features.

the literature to as CANDECOM/PARAFAC, or, in short, CP) [32], [33]. It decomposes a three-way tensor into three matrices:

$$\mathcal{X} = \mathbf{ABC} \tag{16}$$

or

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \tag{17}$$

where $\mathbf{a}_r$, $\mathbf{b}_r$, and $\mathbf{c}_r$, are the r-th columns of the matrices $\mathbf{A}$ $\mathbf{B}$, and $\mathbf{C}$, respectively. In the existence of a strong correlation the original tensor can be approximated by the outer product of the first K column vectors of the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, respectively. These matrices explain, respectively, the rows, columns, and lateral dimension of $\mathcal{X}$ in terms of the unique features uncovered by the decomposition.

Evaluation

Fig. 9 displays the results of a tensor decomposition (CANDECOMP/PARAFAC) on a $13 \times 9 \times 342$ ECG tensor. The row tensor consists of 13 heartbeats, 9 channels, and 342 samples for each channel. The subject was jumping when the measurements were taken. The hidden features estimated for the decomposition were 10. With this assumption, the original tensor could be reconstructed with an accuracy of $91\,\%$ using Equation (17).

Each row in the figure represents a single feature viewed from three perspectives. The column on the left depicts how the feature manifests itself in the heartbeats; the column in the middle depicts which source contributes to the manifestation of the feature; the column on the right depicts the temporal aspect of the feature (i.e., how it manifests itself in the sample). The most interesting feature is the one in the third row (right). The figure is somehow flipped, but it is unmistakably a noise free ECG. The modest swell around sample 100 is the **T** wave whereas the one around sample 325 is the **QRS** wave complex. If we examine the corresponding channels in the middle plot, it is apparent that channels 4, 5, and 6 contribute the most to this wave, which are the three ECG leads. Similarly, the plot on the right shows which heartbeats contribute most. Except the first and the last, all heartbeats contributed to some extent, but the contribution of heartbeats 4 and 11 is the most significant. Compared to SVD (ref. to Fig. 8 (fourth plot, right)), the extraction of the ECG signal with tensor decomposition appears to be by far better.

All the other plots describe artifacts. If we closely examine the middle plots where all the sources agree (i.e., have similar scores), we shall ascertain that the fifth, seventh, and ninth plots are of particular importance. Interestingly, the corresponding plots on the right sides are very similar, from which we can conclude that the plots refer to motion artifacts.

VII. Conclusion

In this paper we reviewed the usefulness of four estimation techniques to model and remove motion artifacts. All of them assume that motion artifacts are linearly added to the useful signals. In ICA, it is assumed that at least one of

the input signals has a non-Gaussian distribution. We have demonstrated that this assumption is difficult to justify. In an Adaptive Filter it is assumed that knowledge of the correlation between the reference signals and the useful signal is available. Often, this is ascertained by measurements taken under ideal conditions. The dimensionality reduction techniques, on the other hand, do not use such prerequisites, examining only the existence of multidimensional correlations to decompose the raw measurement sets into non-overlapping clusters. However, there is no readily available and formal approach to interpret the results, making it difficult to use the results in the implementation of the algorithms. Instead, domain knowledge is required for interpretation. Where domain knowledge is available, dimensionality reduction techniques, particularly, tensor decomposition, can yield reliable and reproducible results.

REFERENCES

[1] J. G. Webster, *Medical instrumentation: application and design*. John Wiley & Sons, 2009.

[2] W. Dargie, *Principles and Applications of Ubiquitous Sensing*. John Wiley & Sons, 2017.

[3] D. Tong, K. Bartels, and K. Honeyager, "Adaptive reduction of motion artifact in the electrocardiogram," in *EMBS/BMES Conference*, Houston, TX, USA, 2002.

[4] M. Raya and L. Sison, "Adaptive noise cancelling of motion artifact in stress ecg signals using accelerometer," in *Engineering in Medicine and Biology*, 2002.

[5] A. Burns, B. R. Greene, M. J. McGrath, T. J. O'Shea, B. Kuris, S. M. Ayer, F. Stroiescu, and V. Cionca, "Shimmer–a wireless sensor platform for noninvasive biomedical research," *IEEE Sensors Journal*, vol. 10, no. 9, pp. 1527–1534, 2010.

[6] Y. Liu and M. G. Pecht, "Reduction of skin stretch induced motion artifacts in electrocardiogram monitoring using adaptive filtering," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE, 2006, pp. 6045–6048.

[7] M. Guermandi, R. Cardu, E. F. Scarselli, and R. Guerrieri, "Active electrode ic for eeg and electrical impedance tomography with continuous monitoring of contact impedance," *IEEE transactions on biomedical circuits and systems*, vol. 9, no. 1, pp. 21–33, 2014.

[8] Y.-J. Huang, C.-Y. Wu, A. M.-K. Wong, and B.-S. Lin, "Novel active comb-shaped dry electrode for eeg measurement in hairy site," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 1, pp. 256–263, 2014.

[9] I. Beraza and I. Romero, "Comparative study of algorithms for ecg segmentation," *Biomedical Signal Processing and Control*, vol. 34, pp. 166–173, 2017.

[10] ShimmerSensing, "Shimmer3 with strap." [Online]. Available: http://www.shimmersensing.com/products/shimmer3-development-kit

[11] J. R. Wessel, "Testing multiple psychological processes for common neural mechanisms using eeg and independent component analysis," *Brain topography*, vol. 31, no. 1, pp. 90–100, 2018.

[12] F. Artoni, A. Delorme, and S. Makeig, "Applying dimension reduction to eeg data by principal component analysis reduces the quality of its subsequent independent component decomposition," *NeuroImage*, vol. 175, pp. 176–187, 2018.

[13] S. Çınar and N. Acır, "A novel system for automatic removal of ocular artefacts in eeg by using outlier detection methods and independent component analysis," *Expert Systems with Applications*, vol. 68, pp. 36–44, 2017.

[14] R. M. Rangayyan, *Biomedical signal analysis*. John Wiley & Sons, 2015, vol. 33.

[15] A. G. Correa, L. L. Orosco, P. Diez, and E. L. Leber, "Adaptive filtering for epileptic event detection in the eeg," *Journal of Medical and Biological Engineering*, vol. 39, no. 6, pp. 912–918, 2019.

[16] L. Murali, D. Chitra, T. Manigandan, and B. Sharanya, "An efficient adaptive filter architecture for improving the seizure detection in eeg signal," *Circuits, Systems, and Signal Processing*, vol. 35, no. 8, pp. 2914–2931, 2016.

[17] R. Martinek, R. Kahankova, H. Nazeran, J. Konecny, J. Jezewski, P. Janku, P. Bilik, J. Zidek, J. Nedoma, and M. Fajkus, "Non-invasive fetal monitoring: A maternal surface ecg electrode placement-based novel approach for optimization of adaptive filter control parameters using the lms and rls algorithms," *Sensors*, vol. 17, no. 5, p. 1154, 2017.

[18] T. Y. Liu, K. J. Lin, and H. C. Wu, "Ecg data encryption then compression using singular value decomposition," *IEEE journal of biomedical and health informatics*, vol. 22, no. 3, pp. 707–713, 2017.

[19] R. Kumar, A. Kumar, and G. K. Singh, "Hybrid method based on singular value decomposition and embedded zero tree wavelet technique for ecg signal compression," *Computer methods and programs in biomedicine*, vol. 129, pp. 135–148, 2016.

[20] W. Dargie and J. Lilienthal, "Application of svd for removing motion artifacts from the measurements of a wireless electrocardiogram," in *2019 22th International Conference on Information Fusion (FUSION)*. IEEE, 2019, pp. 1–8.

[21] H. He and Y. Tan, "Automatic pattern recognition of ecg signals using entropy-based adaptive dimensionality reduction and clustering," *Applied Soft Computing*, vol. 55, pp. 238–252, 2017.

[22] R. J. Martis, U. R. Acharya, K. Mandana, A. K. Ray, and C. Chakraborty, "Application of principal component analysis to ecg signals for automated diagnosis of cardiac health," *Expert Systems with Applications*, vol. 39, no. 14, pp. 11 792–11 800, 2012.

[23] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[24] M. Akhbari, M. Niknazar, C. Jutten, M. B. Shamsollahi, and B. Rivet, "Fetal electrocardiogram r-peak detection using robust tensor decomposition and extended kalman filtering," in *Computing in Cardiology 2013*. IEEE, 2013, pp. 189–192.

[25] S. Padhy and S. Dandapat, "Third-order tensor based analysis of multilead ecg for classification of myocardial infarction," *Biomedical Signal Processing and Control*, vol. 31, pp. 71–78, 2017.

[26] J. Lilienthal and W. Dargie, "Extraction of motion artifacts from the measurements of a wireless electrocardiogram using tensor decomposition," in *2019 22th International Conference on Information Fusion (FUSION)*. IEEE, 2019, pp. 1–8.

[27] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002.

[28] M. Rosenblatt, "A central limit theorem and a strong mixing condition," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 42, no. 1, p. 43, 1956.

[29] J. V. Stone, *Independent component analysis*. Wiley Online Library, 2004.

[30] P. Regalia, *Adaptive IIR filtering in signal processing and control*. Routledge, 2018.

[31] W. Dargie, "Tensor-based resource utilization characterization in a large-scale cloud infrastructure," in *Proceedings of the 12th IEEE/ACM International Conference on Utility and Cloud Computing*, 2019, pp. 83–91.

[32] A.-H. Phan, P. Tichavskỳ, and A. Cichocki, "Fast alternating ls algorithms for high order candecomp/parafac tensor factorizations," *IEEE Transactions on Signal Processing*, vol. 61, no. 19, pp. 4834–4846, 2013.

[33] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—part ii: Definitions and uniqueness," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1033–1066, 2008.