Wavelet-Independent Component Analysis to remove Electrocardiography Contamination in surface Electromyography

Joachim Taelman, Sabine Van Huffel, Senior Member, IEEE, Arthur Spaepen

Abstract—Removing artifacts from biomedical signals has become a major research topic in biomedical signal processing. In electromyography signals a source of contamination is the electrophysiological signal of the heart. This contamination influences the features extracted from the sEMG signals, especially in low-activity measurements of the muscles like stress. As the heart is a muscle, the frequency content of the heart signals overlaps the frequency content of the muscle signals, so basic frequency filtering is not possible. In this paper, we present the results of a recently developed algorithm wavelet-independent component analysis and we compare these results with the widely described algorithm of ECG template subtraction to remove the ECG contamination.

I. INTRODUCTION

Electrophysiological signals are nowadays widely used to derive important information regarding the physiological and pathophysiological behavior of the human body. Electrocardiography (ECG), which measures the electrical signals of the heart, is widely accepted as a method to estimate the physiology and the pathophysiology of a human heart. Since the nineties, electroencephalography (EEG) and electromyography (EMG), which measures the electrical signals of respectively the brains and the muscles, are becoming a field of interest for researchers.

The information about the muscle activity can be retrieved from the signal measured at the skin, the so-called surface EMG (sEMG) or at a muscle-fibre in the muscle itself with a needle electrode. The electrodes record the electrophysiological signal of the muscle. Nerve cells and muscle fibres are depolarized when activated by a certain threshold voltage. The result is the propagation of a depolarization wave along the nerve and muscle fibre. This electrical wave is the direct cause of muscular contraction. A muscle exists of several motor units, a unit of one nerve and the corresponding muscle fibres. The measured sEMG is the sum of the depolarisation waves of several motor units in the environment of the electrode.

As the electrophysiological signal of the muscle contains information, some features can be extracted from the sEMG to predict its behavior. At this moment, the features that are mostly extracted from a sEMG are related to its amplitude (root mean square, RMS) and to its frequency content (mean and median power frequency, MPF and MdPF). These features are very sensitive to artifacts in the sEMG, especially when looking at small signals. In the ConText-project [1], [2], we are trying to infer stress from change in muscle activity. It has been indicated that stress induces an increase in muscle activity, however this increase is subtle. A loss in resolution in the frequency content and amplitude of the signal resulted in less accuracy, so the number of artifacts in the EMG signal must be reduced.

A major source of contamination in a sEMG signal is the electrophysiological signal of the heart (ECG) [3], [4], [5]. As the heart is a muscle, the frequency range of the ECG is similar to this of the EMG. High pass filtering, which is indicated as a possible solution, implies that useful information will be disregarded as it will be filtered out. It has been stated that the ECG-contamination induces huge errors with respect to the amplitude and frequency of the raw signal, especially for low-activity signals.

In the present paper, two algorithms to remove the ECG artifacts are proposed. One of them is based on the algorithm known as template subtraction [3], [4], [5]. This is a simple algorithm where a trained template is subtracted from the measured signal on the heartbeat to remove the ECG-contamination. The second technique is a recently developed algorithm based on wavelet-Independent Component Analysis (wICA) [6]. This algorithm combines a wavelet analysis with a blind source separation technique to overcome the shortcomings of wavelet analysis (overlapping spectrum) and ICA (lack of redundancy of the number of channels compared with the number of sources) separately. The results of these two algorithms are compared.

II. METHODS

A. EMG recordings

The sEMG signals from retrieved from a protocol to study the effect of mental tasks to the activity of the muscles in the neck region. We stated a protocol where four conditions can be distinguished: rest, postural load, mental load and postural load in combination with mental load. It has been indicated that a mental task induces an increase in muscle
activity in the neck muscles [7]. In this test, the M. Trapezius pars descendens, M. Infraspinatus and the M. Deltoideus medius from the left and the right shoulder are measured, together with the heart rate (ECG) of the test person. The sEMG is measured differentially with pre-gelled contact electrodes (Ag-AgCl, 10 mm diameter, Nikomed, Denmark) placed on the positions according to SENIAM [8]. The signals were digitized at a sample rate of 1000 Hz.

In this paper, the signals from one test person (male, 23 years old) are used to show the results of the two described algorithms.

B. Algorithms

1) ECG template subtraction: ECG template subtraction takes the advantage of the periodic characteristics of the ECG signal. This involves subtraction of an ECG template from the EMG signal at each occurrence of the heartbeat. In the measured signals, an advantage is taken from the measured heart rate: the moment heart beat occurs in the measured EMG signals can be localized very accurate. With the Pan-Tompkins algorithm [9], the position of the QRS-complex in the measured ECG signal can be determined very accurately.

The ECG template subtraction algorithm can be divided in two steps. The first step involves the formation of the ECG template contaminating the EMG signal as the shape of the ECG contamination depends on the position and the direction of the electrodes on the muscles. By ensemble averaging a window (500 samples) of EMG signal in the rest state at every heart beat, a template for every channel is trained. This method is based on the assumption that an EMG signal has a zero mean Gaussian distribution. The result of this averaging is a template of the ECG contamination, with a width of 500 samples, which is used as subtraction template [5].

The second step is to subtract the template from the EMG signal to clean them from the ECG contamination. For the detection of an occurrence of contamination, measured heart rate is used, where every heart beat is localized. At each detected heart beat, the template is subtracted from the EMG signal.

The delay between the ECG-signal at the heart and the ECG-contamination in the EMG signal is brought into when the template is trained from the signal itself. These two steps are repeated for every measured channel. ECG template subtraction can only be used on a single channel measurement as the template is different for every channel.

2) Wavelet-Independent Component Analysis (wICA): wICA [6] combines two techniques, wavelet analysis and ICA, to overcome the shortcomings of the two techniques separately.

- The wavelet transform [10] is a time-frequency representation of a signal that was introduced to overcome the limitations in time and in frequency resolution suffered by the classical Fourier and their evolutions like Short Time Fourier Transform (STFT). In stead of using a sine wave as basic function, like the Fourier transform, a basic waveform is used. This basic waveform $\psi$ can be modified to basic functions $\psi_{a,b}$ obtained from dilations and shifts of the basic waveform. This basic waveform is shown in (1), where $a$ is the scaling parameter and $b$ represents the translation parameter.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi \left( \frac{t-b}{a} \right)$$ (1)

The continuous wavelet transform (CWT) of a given input signal $x(t)$ is shown in (2).

$$CWT(a,b) = \int x(t)\psi_{a,b}^*(t)dt$$ (2)

The representation of the CWT is redundant. The discrete wavelet transform (DWT) removes the redundancy of the CWT by using discrete steps for scale and translation.

The DWT can be used to perform artifact removal. Its application is based on a good spectral separation between the original signal and the artifact. The DWT must be applied to a single channel recording as it separates this single channel in multiple channels.

- Independent Component Analysis (ICA) [11] is a method for solving the Blind Source Separation problem. This technique extracts statistically independent components $S$ form a set of measured signals $X$ (3).

$$X = AS$$ (3)

The goal of ICA is to estimate the mixing matrix $A$ and/or the source vector $S$ from the measured data $X$. ICA makes the assumption that the different sources are statistically independent and uses this assumption to estimate the mixing matrix recursively.

Unlike wavelet analysis, ICA can only be applied to multichannel recordings. As we are working with real data, the number of recorded channels is not enough to retrieve the real sources, hidden in the data.

- Wavelet-Independent Component Analysis combines these two techniques. In the first step, a wavelet analysis is applied on each channel. Each channel is split up in a number of wavelet components, which will be used as input for the ICA. With this improvement, more information of the original input signal will be used to find the independent sources.

The second step of this algorithm is to find the original data, cleaned from the artifacts. When a source of contamination is found in one component of the ICA, this source can be set to zero. From the inverse step of the ICA, we are able to find the wavelet-components. From these components, the original cleaned up signal can be reconstructed.

C. Validation

The validation which is used to compare the results is the visual validation. The results from the cleaning algorithms
are compared with the raw signal. As the algorithm of template subtraction is widely described and used, these results will be used as reference to validate the results of the recently developed wICA algorithm.

III. RESULTS

In this section, the results of the algorithms are shown. Figure 1 shows the raw data. The contamination of the heart beat in the measured sEMG signals is clear in channel two, four and six. In channel three and five, traces of the contamination can be found. From figure 1, the influence from the electrical signals of the heart to the electrical signals from the muscle is evident.

The first algorithm to clean the EMG signals, described in this paper, is the template subtracting algorithm. As mentioned before, this algorithm is performed on a single channel. From every measured muscle, a subtraction template is trained. At every heart beat, detected in the measured ECG signal, the template is subtracted from the correspondent EMG. The results of this action are shown in figure 2. All the ECG artifacts, which are clearly visible in figure 1, are cleaned up with this algorithm.

The second algorithm, described in section II, is the use of wICA to eliminate the ECG contamination. Before the algorithm can be applied, some preliminary decisions must be taken. The first choice is the inclusion of the measured ECG signal to the system. The test has been done with and without the measured heart beat. The conclusion was drawn that the algorithm performs better when the measured ECG signal was included: The heart beat and the heart beat contamination on the other channels became the dominant sources and the separation in sources was more accurate. As the ECG signals are measured during the test, the decision was taken to include the ECG signals as data.

The second choice was the decision of the DWT. The ‘Daubechies 6’ [10] was used as basic wavelet and six decompositions performed well.

The results of applying this algorithm to the original data resulted in six dominant ECG sources and eight other relevant sources. The reconstruction of the cleaned signal is shown in figure 3. When the reconstructed signal is compared with the signals after template subtraction (which has been stated as the reference), it is clear that the algorithm performs well on four of the six channels (channels 1, 3, 4 and 6). The ECG contamination is also reduced in the reduced in the reconstruction of the other two (the measurements on the M. Infraspinatus right and left) but the reconstructed EMG signals are somewhat different compared to the reference. This indicates that the wICA algorithm was not able to find the real sources of these two signals.

An improvement was performed when splitting up the original signal in two data sets: first test set comprises
original channel 1, 3, 4 and 6 while the second data set contains the other two channels. The wICA algorithm is performed separately on these two data sets. The ECG signal was included in each of the two data sets. The results of this test are shown in figure 4. The reconstructions of the data on channels 1, 3, 4 and 6 are similar to the results of the whole data set. However, the reconstructions channels 2 and 5 are better, comparing to the results of figure 3. This indicated that the sources of channels 1, 3, 4 and 6 dominate the sources of channels 2 and 5.

It can be concluded that this wICA algorithm is able to clean the sEMG from its ECG contamination.

**IV. DISCUSSION**

In this paper, the results of a recently developed algorithm wICA to eliminate artifacts in biomedical signals are used to eliminate ECG contamination. The results were compared with a known and robust template subtraction algorithm. The results are promising as this algorithm was able to eliminate the ECG artifacts. The algorithm performed even better when the original data of six channels was split up in a logical way in two data sets of two and four channels. A drawback of the algorithm is that it works only on multichannel data.

It must be stated once more that the measured ECG signal is included as raw data. This outperformed the wICA when the measured ECG signal was not included; however, the algorithm without the ECG signal was able to improve the amount of contamination in the reconstructed data.

The decision on the wavelet type and the number of decomposition was taken arbitrary. Further research will be needed to come to a good conclusion about the use of the wavelet type and the number of decompositions.

**ACKNOWLEDGMENT**

We thank the European partners of ConText: Philips Research (The Netherlands), TNO (The Netherlands), Clothing + (Finland), TITV (Germany), TU Berlin (Germany)

**REFERENCES**


