PCA-based Noise Reduction in Ambulatory ECGs

I Romero

IMEC, Eindhoven, the Netherlands

Abstract

PCA can be used for cleaning noisy ECGs. With this aim, ECG with artificial motion artifacts were generated by combining clean 8-channel ECG with noise signals. 8-channel PCA was applied and then inverted after selecting a subset of principal components (PC). Input and output of PCA filtering was compared by calculating the correlation coefficient and estimating the SNR.

Above 0dB, the PC corresponding to highest variance gave best performance, below 0dB the best PC was the second highest or lower variance. When SNR decreased, PCA performed better when retaining more number of PCs (3 PCs for a SNR=10dB down to 6 out of 8 PC for SNR=-10dB). Reducing the number of input ECG channels did not yield to a significant difference when it was reduced from eight down to two. A method for identifying the optimal subset of PC as a function of input SNR and number of channels was proposed. This method achieved an SNR improvement of 0.95dB-1.92dB.

1. Introduction

During the last 20 years new ambulatory cardiac monitors have been developed for continuous ECG monitoring. These devices are portable with an autonomy which is increasing with the improvement of low-power micro-electronics. Integration of microprocessors allows performing some signal processing and automatic interpretation. However, in ambulatory conditions, noise increases with higher levels of activity of the subject (e.g. during exercise). With movement, motion artifact energy could be even higher than that of the ECG, resulting in a reduction of signal quality that would make interpretation very difficult.

Several methods for noise reduction and motion artifact removal have been proposed in literature. Traditional denoising techniques were based on time averaging [1] and frequency analysis such as filter banks [1] or the wavelet transform [2]. In adaptive filtering, a filter is applied after adjusting its parameters in time to a time varying noise. This is particularly useful when the noise is non-stationary as is the case in ambulatory motion artifacts. However, a reference signal has to be additionally recorded together with the ECG. Several adaptive filtering approaches have been proposed to obtain an adequate reference signal such as measurement of skin-electrode impedance [3, 4], skin stretching measured with optical sensors [4, 5] or accelerometers [6, 7].

As sources of ECG and motion artifacts are uncorrelated, blind source separation (BSS) techniques could be used for separating both signals [8, 9]. In order to apply those methods, a multi lead ECG recording is required and the different recorded leads should be linearly independent. Although these conditions are commonly met in ambulatory holter monitors, there is very little in literature describing the use of BSS techniques for ECG denoising. Principal Components Analysis (PCA) has been used for reducing noise in single lead ECG segmented in time intervals [10]. A combination of PCA and ICA was also proposed by Chawla [11] for ECG denoising.

For ambulatory applications, it should be taken into account the limitations in computational power and memory. With this aim, PCA was investigated for its relatively low computational complexity.

2. Methods

2.1. Principal component analysis (PCA)

Principal Component Analysis (PCA) is a technique this is commonly used in multivariate statistical analysis. Its goal is the reduction in the number of dimensions from a numerical measurement of several variables. With this dimensional reduction, this technique looks for simplifying a statistical problem with the minimal lost of information. This method is also used in signal processing for separating a linear combination of signals generated from sources that are statistically independent. This is performed by representing the data with a new coordinate system. This transformation is bidirectional and no information is lost [12].

Applying PCA to n ECG leads that are statistically independent gives n new signals or principal components. The first signal corresponds to the principal component with highest variance while the *n-th* signal corresponds to the principal component with the lowest variance. In low noise conditions, principal components with higher variances have information mainly from the ECG while low variance's components correspond to noise [13]. However, in ambulatory conditions, motion artifacts can have higher energy levels than the ECG signal, making selection of principal components more complex.

2.2. Data collection

Clean ECG signals were obtained by recording 8-unipolar lead ECGs from 5 healthy subjects. For each subject 8 sets of 10 seconds were obtained while the subject was at rest. 8-channel noise recordings were obtained by placing 9 electrodes on the back of the subjects at the height of the lumbar curve where ECG signals were negligible. Then, the subjects were asked to move randomly. For each subject, 8 sets of 10 seconds were recorded. Signals were obtained with a sampling frequency of 1000 Hz, using a generic biosignal acquisition system from g.Tec (g.USBamp). All recordings were filtered by a high pass filter with cut-off frequency of 0.5 Hz and a 50 Hz notch filter. Each 8-channel noise signal was multiplied for a gain factor and added to each 8-channel clean ECG in order to obtain a specific SNR. SNR values ranging from 10 to -10 dB were considered. For each SNR value 400 different combinations of clean ECG and noise were selected. Figure 1 shows one example of a clean ECG, a pure noise and a combination of both signals.

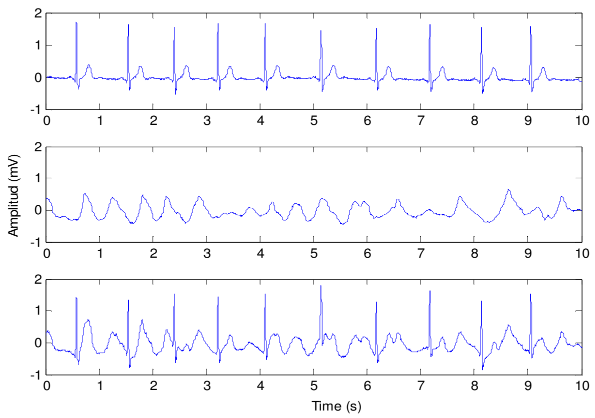

Figure 1. Extract of a clean ECG, a noise and the sum of both signals. The SNR of the combined signal is of 0 dB.

2.3. Evaluation criteria

PCA was applied and then inverted after selecting a subset of principal components. For evaluating the signal improvement, the correlation coefficient between the noise-free signal and the output after PCA filtering was computed. In addition, Signal to Noise Ratio (SNR) before and after PCA filtering was estimated. In order to calculate the SNR, the noise energy was estimated by subtracting a template averaged beat at the location of every R peak. The median and median absolute deviation (MAD) values over all ECG leads were considered as representative values for each signal. As median and MAD are more robust to outlier values, they were preferred to the arithmetic mean and standard deviation.

3. Results

Half of the dataset (i.e. 200 signals for each SNR value) with simulated signals formed with the combination of clean ECG and motion artifact noise at different SNR values was used as a *learning dataset* to evaluate the performance of Principal Component Analysis (PCA).

SNR values from 10 to -10 dB in steps of 1 dB were considered. Both the noisy signal and the output signal from PCA filtered were then compared with the clean ECG in order to study the improvement due to the algorithm.

3.1. Best principal component

Initially, PCA was applied to the 8 ECG lead signals. Only one of the 8 resulting components was retained and the PCA projection was inverted in order to obtain the filtered original 8 ECG leads.

Selecting the principal component (PC) with highest variance gave in general highest correlation coefficients for high SNR (over 0 dB). However, for SNR values between 0 and - 7dB the principal component which had highest ECG content was the second one with highest variance. Between -8 and -10 dB the PC which gave the best the highest correlation coefficient was the 4^{th} one.

Overall, correlation coefficient values were below the correlation values between noisy and clean ECG. The correlation coefficient dropped in median 0.08 (MAD = 0.09) and the SNR improvement was in median of 0.81 dB (MAD = 2.34) when only the component with highest variance was retained.

3.2. Number of principal components

The optimal number of principal components in function of the SNR was then investigated. PCA was applied to the 8 ECG lead signals. Then, the PC's were sorted by their ranking obtained as explained in previous section (i.e. correlation coefficient) in descending order. Finally, *n* (*n*=1, 2, ..., 8) components were selected and the PCA projection was inverted.

Overall, the best values were obtained with n=3 giving a small correlation coefficient improvement with median of 0.02 (MAD = 0.05) and the SNR improvement was in median of 1.49 dB (MAD = 4.62). For low SNR's, PCA performed better when retaining more number of PCs (4 PC's for a SNR=-8dB and 6 PC for SNR=-9 and -10dB).

3.3. Principal component selection as a function of SNR

Combining the results obtained in the previous sections, the best subset of PC's was selected for each SNR value. Figure 2 shows the subset of PCs that gave best results when considering correlation.

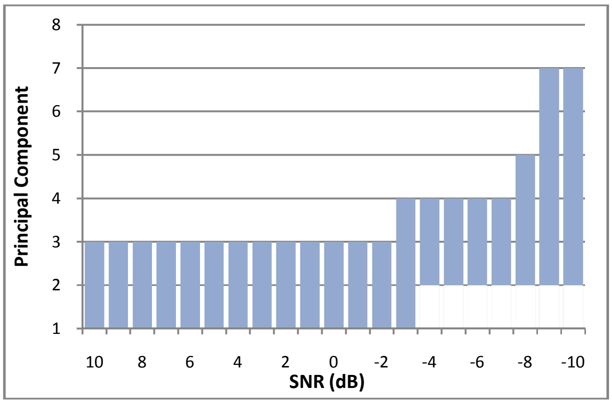

Figure 2. Subset of PCs as a function of the SNR. PC 1 corresponds to the one with highest variance. PC 8 corresponds to the one with lowest variance.

In order to evaluate the performance of a PCA-based noise reduction algorithm with this PC's selection, the second half of the dataset was considered (200 signals) as *evaluation dataset*. PCA was computed, the best PC subset for each SNR values (as plotted in figure 1) was retained and PCA projection was inverted.

As a first reference, the results were compared with the direct comparison of the noisy ECG and the clean ECG (i.e. no denoising).

As a second reference the optimal set of principal components was calculated for each signal within the dataset and SNR value independently. The optimal set was defined as the one (from all possible combinations) that gave the highest correlation coefficient between the clean ECG signal and the output of the inverted PCA.

The median correlation coefficient for each SNR value and method is plotted in figure 3. Figure 4 shows the median SNR improvement for each SNR value. As can be seen in the figures, applying PCA can give a significant improvement, especially with low SNR values (improvement in the correlation coefficient of 0.16 and SNR 6.39 dB with SNR=-10dB) when the optimal PC can be indentified for each noisy signal. However, the subset of PC's as a function of SNR as proposed above, gave a smaller improvement (improvement in the correlation coefficient of 0.03 and SNR 1.92 dB with SNR=-10dB).

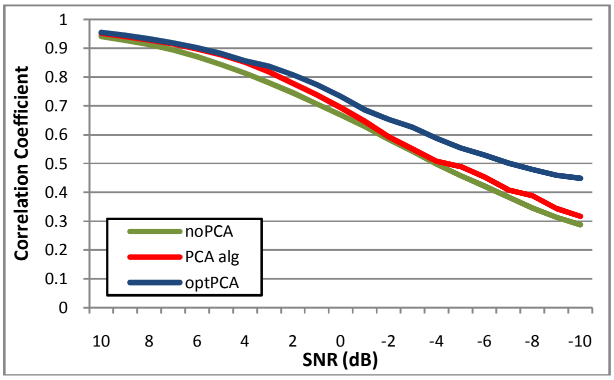

Figure 3. Correlation coefficient of PCA output and clean ECG signals plotted against SNR value when the selected PC were retained (PCA alg), the optimal subset of principal components were retained (optPCA) and the median correlation coefficient of noisy ECG with clean ECG signals (noPCA).

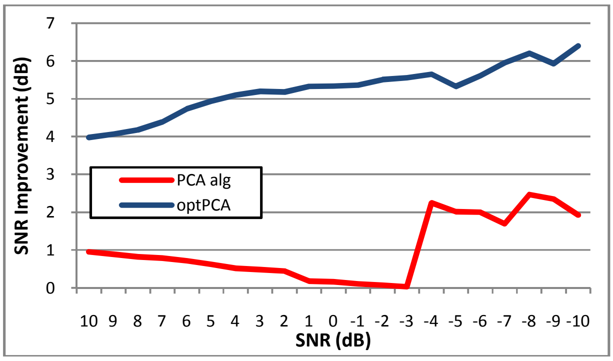

Figure 4. SNR improvement of PCA output and clean ECG signals plotted against SNR value when the selected PC were retained (PCA alg) and the optimal subset of principal components were retained (optPCA).

3.4. Number of input channels

The number of input channels was also investigated. In addition to eight input channels, subsets of input channels of six, four and two channels were also considered.

Following the same procedure as in the previous study, for each input subset, the optimal component subset (considering the highest correlation coefficient) was found for each signal and SNR value. To obtain consistent results only one lead (common for all subsets) was considered for comparing the input and output of the PCA algorithm.

Overall, the improvement in the correlation coefficient was very similar for all cases. The best results were obtained when used eight input leads (median=0.78 MAD=0.16). For six (median=0.71, MAD=0.17), four (median=0.72, MAD=0.17) and two (median=0.74, MAD=0.19) input leads the correlation coefficient improved slightly less.

The SNR improvement gave bigger differences with median of 4.35 dB (MAD=3.37) for eight input leads, 3.95 dB (MAD=3.46) for six, 2.72 (MAD=2.94) for four and 0.38 (MAD=3.66) for two. Values are given for the optimal component subset.

Figure 5 shows the results against the different values of SNR considered. Decreasing the number of input signals did not yield to a big drop in the correlation coefficient (median drop of 0.49 and 0.42 for 8 and 2 input leads respectively at SNR=-10dB). The SNR improvement dropped in median value from 4.35 dB down to 0.38 dB (8 and 2 input leads) at SNR=-10dB.

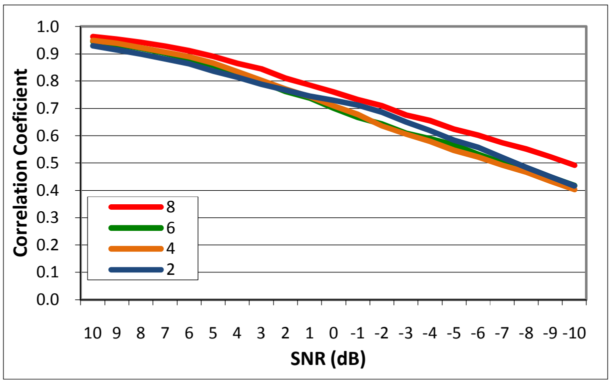

Figure 5. Correlation Coefficient of PCA output and clean ECG signals plotted against SNR value for different number of input signals. The optimal principal component for each signal and SNR was retained.

4. Conclusions

This work investigated the performance of PCA in denoising ECG signals recorded in ambulatory conditions. A simulated database formed by the combination of clean ECG signals with noise scaled to different levels of energy was developed for evaluation.

It was observed, that for high SNR values, retaining the principal components (PC) of highest variances gave best performance. When SNR decreased, the PCs corresponding to highest variance were related to high amplitude noise. Reducing the number of input ECG channels did not yield to a big difference when it was reduced from eight down to two.

A method for identifying the optimal subset of PC as a function of input SNR and number of channels was proposed. This method achieved an SNR improvement of 0.95 dB at 10 dB of SNR and 1.92 dB at SRN=-10dB.

As a limitation of this study, it should be noted that some stationary has been assumed as signals were of 10 seconds length. The performance under shorter duration noise was not studied. It should be noted that the parameters used for evaluation the signal improvement were the correlation coefficient and the estimation of SNR which for some applications, they may not always characterize best the signals under study.

It was observed that an optimal selection of PCs yielded to a significant improvement of SNR. This suggests that other method for automatic component selection could lead to a better performance. Furthermore, other techniques for signal decomposition such as Independent Component Analysis could yield to a better separation of signal and noise, although that would increase the computational complexity.

References

[1] Afonso V, Tompkins W, Nguyen T, Michler K, Luo S. Comparing stress ECG enhancement algorithms. IEEE Eng Med Biol Mag 1996;15:37–44.

[2] Augustyniak P. Separating cardiac and muscular ECG components using adaptive modelling in time-frequency domain. Proc. of the WACBE World Congress on Bioengineering 2007.

[3] Devlin PH, Mark RG, Ketchum JW. Detection electrode motion noise in ecg signals by monitoring electrode impedance. Computers in Cardiology 1984; 51-56.

[4] Hamilton P, Curley M, Aimi R, Sae-Hau C. Comparison of methods for adaptive removal of motion artifact. Computers in Cardiology 2000; 27: 383– 386.

[5] Hamilton P, Curley M, Aimi R. Effect of adaptive motion-artifact reduction on QRS detection. Biomed Instrum Technol 2000; 34: 197–202.

[6] Tong D, Bartels K, Honeyager K. Adaptive reduction of motion artifact in the electrocardiogram. Proc. Second Joint EMBS/BMES Conf 2002; 2: 1403–1404.

[7] Raya M, Sison L. Adaptive noise cancelling of motion artifact in stress ECG signals using accelerometer. Proc. Second Joint EMBS/BMES Conf 2002; 2: 1756–1757.

[8] Hyvärinen A, Oja E. Independent component analysis: Algorithms and applications. Neural Networks 2000; 13: 411–430.

[9] Castells F, Cebrián A, Millet J. The role of independent component analysis in the signal processing of ECG recordings. Biomed Tech (Berl) 2007; 52(1): 18–24.

[10] Palaniappan R, Khoon TE. Uni-channel PCA for noise reduction from ECG signals. Proc 1st International Bioengineering Conference 2004; 436-439.

[11] Chawla MPS. A comparative analysis of principal component and independent component techniques for electrocardiograms. Neural Comput & Applic 2009; 18: 539-559.

[12] Jolliffe IT. Principal Component Analysis. Springer Series in Statistics 2002. New York: Springer.

[13] Castells F, Laguna P, Sornmo L, Bollmann A, Millet-Roig J. Principal Component Analysis in ECG Signal Processing. EURASIP J. Appl. Signal Process 2007; 2007(1): 98–119.

Address for correspondence.

Iñaki Romero.
High Tech Campus 31
5605 KN Eindhoven
The Netherlands
romero@imec-nl.nl