Component Selection for Principal Component Analysis-Based Extraction of Atrial Fibrillation

I Romero Legarreta

Physikalisch-Technische Bundesanstalt, Berlin, Germany

Abstract

For the study of atrial fibrillation (AF) in the surface ECG, the cancellation of the QRS-T is required in order to isolate the atrial from the ventricular activity. Principal Component Analysis (PCA) was previously employed with good results. The main problem with this method is the selection of the principal components that contains the AF wave information. This paper presents a study to determine the best subset of the 12 principal components computed from a 12 lead standard surface ECG in order to optimize performance.

A test database consisting of 840 ECGs with simulated AF was developed. This test dataset was used to determine the performance of the PCA when retaining different subsets of the principal components. It was observed that the components 3 to 8 contributed mainly to the atrial fibrillation wave.

Finally, the best PCA variant found was used to analyse the PTB AF database. The distribution of the main frequencies and the concentration of the spectral energy around the main frequencies were determined for this database..

1. Introduction

Atrial Fibrillation (AF) is probably the most frequent cardiac arrhythmia. The treatment of this pathology is complex and usually based on trial and error; treatment results are still unsatisfactory and unpredictable. For this reason the study of AF has received increasing interest in the scientific community during the last years.

The use of new digital processing techniques has permitted to extract the atrial activity for further analysis by proper cancellation of the ventricular activity [1]. For this purpose, different promising techniques have been proposed recently, e.g. spatiotemporal QRST cancellation (STC) by subtraction [2,3], application of Independent Component Analysis (ICA) [4,5] or Principal Component Analysis (PCA) [6,7]. A comparison study of these three main techniques: STC, ICA and PCA led to satisfactory results for this three methods [10]. However, the main difficulty in the application of ICA and PCA is the selection of the components that best match with the atrial source signal. Application of these cancellation methods permits the study of the atrial fibrillation wave in the frequency or time-frequency domain [8,9].

After the extraction of the AF wave, the frequency analysis gives a better understanding of the phenomena involved in atrial fibrillation and its behaviour [9,11]. In addition, it can be utilized by cardiologists for diagnosis and therapy control [12,13].

2. Methods

A. Test AF signals

For the developing and testing of an AF wave extraction method, a dataset of ECGs with simulated AF was created. An AF wave was simulated from a real AF ECG by manually cancelling the segments corresponding to QRS-T waves. The gaps resulted were filled with the interpolation of adjacent AF segments. A QRS-T wave was obtained from real ECGs in sinus rhythm by manually cancelling the P waves. The test signals were created by the addition of both AF and QRST waves. One example is shown in figure 1.

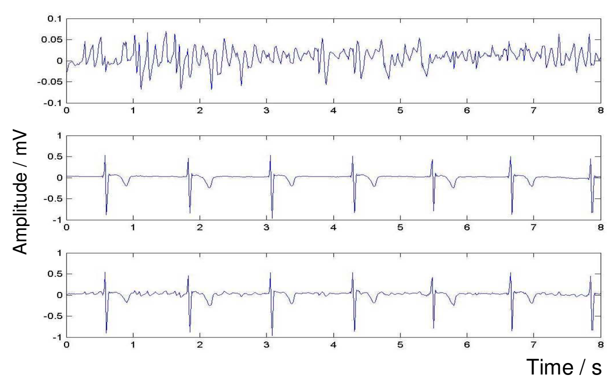

Figure 1. Generation of a Test AF signal. The upper plot shows a simulated AF signal. The middle plot contains a QRS-T signal. The addition of both signals is showed in the lower plot.

Thirty AF waves were simulated from selected real cases within the PTB AF database. Twenty eight QRS-T waves were obtained from real cases within the PTB ECG database. The combination of them gave a total of 840 test signals. Each signal is comprised of the 12 standard leads with 10 seconds length. The sample frequency was of 100 Hz to match the sample rate in the PTB AF database.

B. PTB AF database

The Physikalisch-Technische Bundesanstalt (PTB) ECG database consists of more than 25.000 validated ECGs [14] with measurement times from 10 to 108 seconds. Within the ECG database 546 cases classified as atrial fibrillation have been selected. After additional confirmation by experienced cardiologists four of these cases were removed yielding altogether 542 cases for the AFIB test database. The data were recorded with different sampling frequencies ranging from 400 Hz to 10 kHz. For the AFIB test database the data were downsampled (after low-pass filtering) resulting in a uniform sampling frequency of 100 Hz. Additionally, the length of all ECGs in the AFIB test database was restricted to 10 seconds duration.

C. Spatial Filtering and Component Selection

The test dataset created was used to determine the performance of the Principal Component Analysis (PCA). PCA was applied to the signals. One single component was considered and the inverse PCA was applied in order to obtain the 12 standard leads. Both the estimated and the simulated AF waves were compared by means of the root mean squared error and the correlation coefficient. Only lead V1 was considered because it usually shows the best representation of the atrial activity [9]. The median value of all the correlation coefficient and square error values obtained for the whole test dataset was calculated. These values are represented in figure 2.

As can be seen in figure 2, the principal component 6 obtained the highest correlation index for the test dataset constructed. In fact the components can be resorted considering the median correlation index obtained from the whole test dataset as: 6 5 7 3 4 8 2 1 10 9 11 12. The squared error was sensibly higher for the two first components than for the rest of the components for which the values were very similar.

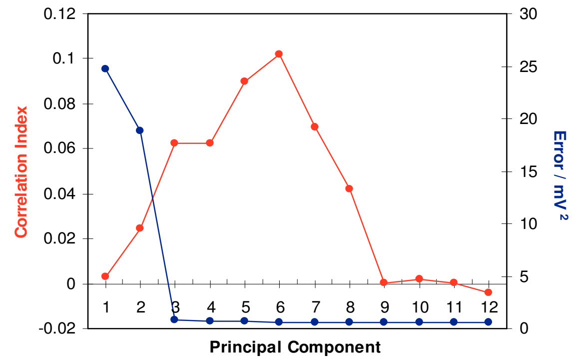

Figure 2. Median values of the correlation and squared error for single principal components in the test AF database.

Considering this new order a further study was carried out by adding components in that order (i.e. retaining component 6, components 6 and 5, components 6, 5 and 7 …). Again, the inverse PCA was applied in order to obtain the 12 standard leads and the estimated and the simulated AF waves were compared by means of the root mean squared error and the correlation coefficient. The median value of all the correlation coefficient and square error values obtained for the whole test dataset was again calculated. These values are represented in figure 3.

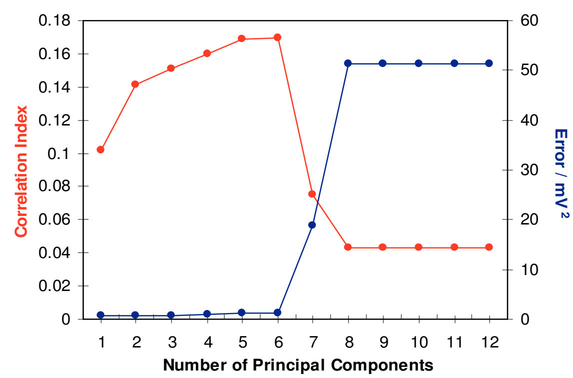

Figure 3. Median values of the correlation and squared error for the addition of principal components in the test AF database.

As can be seen in figure 3, the correlation index increased with the number of components up to 6 components. After that, adding more components resulted in a lower correlation index. The error gave similar results, it was almost constant when adding components up to 6, increasing significantly when adding more components. Therefore the method performed best when considered only the first 6 components in the order discussed above (3 to 8) Figure 4 shows one example of the spatial filtering method.

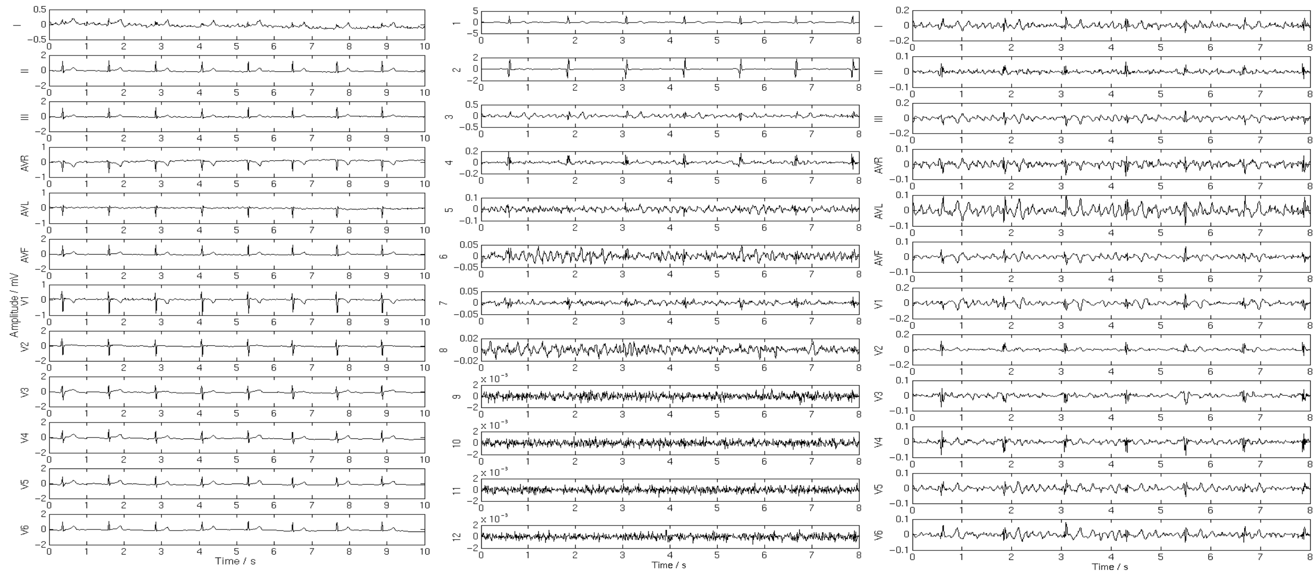

Figure 4. Spatial Filtering on a 12 standard leads ECG. Left figure shows an 12 standard leads ECG. The middle figure corresponds to the 12 Principal Components. Finally, the principal components 3 to 8 are inverted into the original 12 standard leads obtaining the corresponding original ECG filtered.

For assessing the method the inverse PCA was considered in order to obtain the 12 standard leads after retaining a subset of components (spatial filtering). However, the inverse PCA is usually not required since the AF wave is directly obtained by the combination of a subset of components. As can be seen in figure 4 components 1 and 2 corresponds mainly to the ventricular activity while 9 to 10 are mainly noise. Figure 5 shows the combination of principal components 3 to 8.

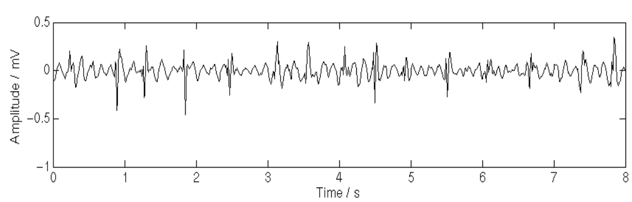

Figure 5. Combination of the components 3 to 8 showed in figure 3.

3. Results

After selected the best subset of principal components for the extraction of the AF wave, the method was applied to real cases from the PTB AF database. The spectrum was computed applying the Fourier Transform and the frequencies of the three main peaks were identified (figure 6). The relative spectral concentration within a 2 Hz window centered around each peak was also calculated (i.e. energy within the window divided by total energy).

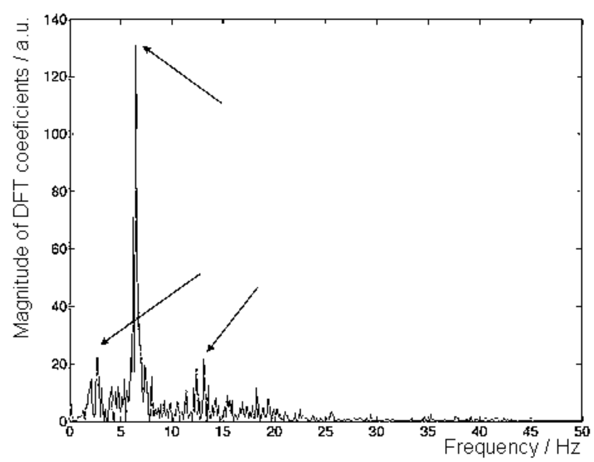

Figure 6. Fourier Transform analysis of an AF wave. The three main peaks are indicated.

The mean (standard deviation) of the frequencies corresponding to the main AF peak calculated from the PTB AF database was 6.24 (4.22)Hz. For the second and third peaks, the results obtained were 7.41 (5.84)Hz and 8.46 (6.79)Hz respectively. The distributions of the three peaks showed that the frequency was mainly contained in an interval of 4 to 9 Hz. The histogram of the first peak's frequencies can be seen in figure 7.

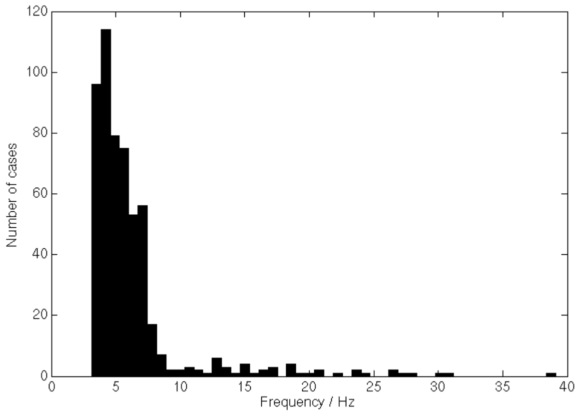

Figure 7. Histogram showing the distribution of the mean peak frequencies.

The spectral concentration gives an idea of how the energy is concentrated around the peaks. The first peak has a ratio of concentration of 0.27 (0.17). The second and third peaks have lower values of concentration with 0.17 (0.09) and 0.13 (0.07).

4. Discussion and conclusions

In the work presented in this paper, the Principal Component Analysis was assessed for extracting the AF wave with a test dataset created by the author. Different subsets of principal components where considered obtaining best results when retaining components 3 to 8 and discarding the rest.

The method was then used to extract the AF wave from real ECGs within the PTB AF database. A spectral analysis of the atrial activity showed that the main frequency peaks were within a range of 4 to 9 Hz and the relative spectral concentration was in average (standard deviation) of 0.27 (0.17) around the mean peak.

Acknowledgements

The author thanks Dr. Clemens Elster and Dr. Gerd Wübbeler for their technical support, and in addition, Prof. Andreas Bollmann and Dr. Daniela Husser for the careful re-validation of the AFIB test data.

This work was supported by the Deutscher Akademischer Austausch Dienst (DAAD).

References

[1] Romero Legarreta I, Xie M, Link A, Bousseljot RD, Kreiseler D, Elster C, Bär M, Koch H. Analysis of the Atrial Fibrillation Main Frequency in the PTB ECG database. ESGCO proceedings 2006: 120-122.
[2] Stridh M, Sörnmo L. Spatiotemporal QRST Cancellation Techniques for Analysis of Atrial Fibrillation: Methods and Performance. Computers in Cardiology 1998; 25:633-636.
[3] Stridh M, Sörnmo L. Spatiotemporal QRST Cancellation Techniques for Analysis of Atrial Fibrillation. IEEE Transactions on Biomedical Engineering 2001; 48(1):150-111.
[4] Rieta JJ, Zarzoso V, Millet Roig J, García Civera R, Ruiz Granell R. Atrial Activity Extraction Based on Blind Source Separation as an Alternative to QRST Cancellation for Atrial Fibrillation Analysis. Computers in Cardiology 2000; 27:69-72.
[5] Steinhoff U. Signal Identification and Noise Suppression in Multi-Channel ECG and MCG by Independent Component Analysis (ICA). In: De Ambroggi L, Katila T, Maniewski R. High Resolution ECG and MCG mapping 2003: Warsaw: International Centre of Biocybernetics: 117-125.
[6] Langley P, Bourke JP, Murray A. Frequency Analysis of Atrial Fibrillation. Computers in Cardiology 2000; 27: 65-68.
[7] Raine D, Langley P, Murray A, Furniss SS, Bourke JP. Surface Atrial Frequency Analysis in Patients with Atrial Fibrillation. Journal of Cardiovascular Electrophysiology 2005; 16(8): 838-844.
[8] Langley P, Stridh M, Rieta JJ, Sörnmo L, Millet-Roig J, Murray A. Comparison of Atrial Rhythm Extraction Techniques for the Estimation of the Main Atrial Frequency from the 12-lead Electrocardiogram in Atrial Fibrillation. Computers in Cardiology 2002;29:29-32.
[9] Stridh M, Sörnmo L, Meurling C, Olsson B. Frequency trends of atrial fibrillation using the surface ECG. Proc. EMBS, IEEE Engineering in Medicine and Biology Society, Atlanta, USA, 1999..
[10] Langley P, Rieta JJ, Stridh M, Millet-Roig J, Sörnmo L, Murray A. Comparison of Atrial Signal Extraction Algorithm in 12-Lead ECGs with Atrial Fibrillation. IEEE Transactions on Biomedical Engineering 2006; 53(2): 343-346.
[11] Langley P, Steinhoff U, Trahms L, Oeff M, Murray A. Analysis of Spatial Variation in the Atrial Fibrillation Frequency from the Multi-channel Magnetocardiogram. Computers in Cardiology 2003; 30: 137-140.
[12] Husser D, Stridh M, Sörnmo L, Olsson B, Bollmann A. Frequency Analysis of Atrial Fibrillation From the Surface Electrocardiogram. Indian Pacing and Electrophysiology Journal 2004; 4(3): 122-136.
[13] Bollmann A, Kanuru NK, McTeague KK, Walter PF, DeLurgio DB, Langberg JJ. Frequency Analysis of Human Atrial Fibrillation Using the Surface Electrocardiogram and Its Response to Ibutilide. American Journal of Cardiology 1998; 81: 1439-1445.
[14] Bousseljot R, Grieger U, Kreiseler D, Schmitz L. Internet-based ECG-Evaluation and Follow-up. 2nd OpenECG Workshop, Berlin, 2004; 53-54.

Address for correspondence

Iñaki Romero Legarreta
PTB – AG 8.42
Abbestr. 2-12
10587 Berlin (Germany)
inaki.romero@ptb.de