Reliable Atrial Activity Extraction from ECG Atrial Fibrillation Signals

Felipe Donoso[1], Eduardo Lecannelier[2], Esteban Pino[1], and Alejandro Rojas[1]

[1] Department of Electrical Engineering, University of Concepcion, Concepcion, Chile
{felipedonoso,estebanpino,arojasn}@udec.cl
[2] Department of Internal Medicine, University of Concepcion, Concepcion, Chile
elecanne@udec.cl

Abstract. Atrial fibrillation (AF) is the most common arrhythmia encountered in clinical research, with a prevalence of 0.4% to 1% of the population. Therefore, the study of AF is an important research field that can provide great treatment improvements. In this paper we apply independent component analysis to a 12-lead electrocardiogram, for which we obtain a 12-source set. We apply to this set three different atrial activity (AA) selection methods based on: kurtosis, correlation of the sources with lead V1, and spectral analysis. We then propose a reliable AA extraction based on the consensus between the three methods in order to reduce the effect of anatomical and physiological variabilities. The extracted AA signal will be used in a future stage for AF classification.

Keywords: atrial fibrillation, atrial activity, ECG, ICA, kurtosis, correlation, power spectral density.

1 Introduction

Atrial fibrillation (AF) is the most common arrhythmia encountered in clinic research, with a prevalence of 0.4% to 1% of the population. This prevalence increases with age, reaching up to 8% in population over 80 years old [1,2]. Therefore, the study of AF is an important research field that can provide great treatment improvements, such as lower morbidity and mortality, better life quality, and lower costs for the health care provider.

AF is characterized by uncoordinated atrial activation with consequent deterioration of atrial mechanical function [1]. From a clinical point of view, the analysis of these activities using non-invasive measures is highly desirable. To this end, the standard 12-lead electrocardiogram (ECG) can be used. Unfortunately, atrial activity (AA) is coupled with ventricular activity (VA), represented by the QRST complex, in all the 12-lead measures. Furthermore, AA presents much lower amplitude than VA, sometimes with amplitudes near the noise level. Additionally, both activities have spectral distributions that overlap, making linear filtering solutions not useful. Hence, one of the most important tasks for an appropriate AF analysis is the dissociation of AA from VA. When using 12-lead ECG the challenge we face for the dissociation of both activities is to

find a representative waveform that estimates the AA, starting from the 12-lead ECG. Some of the algorithms used for AA extraction include spatio-temporal QRST cancellation [3,4], principal components analysis (PCA) [5] and independent component analysis (ICA) [6,7,8].

In this work we apply ICA together with three methods to select the best representation of AA. The goal is to characterize the AA in AF, in order to find patterns that can be recognized as different ECG-based AF classes, in the hope that classes will give complementary information to improve the different treatments currently applied to restore sinus rhythm to patients with this cardiac alteration.

2   Methods

2.1   Data

Real 12-lead ECG recordings were used from four patients with diagnosed AF. Records were 60 s long, sampled at 1200 Hz. A pre-processing consisting in a bandpass filtering and an amplitude normalization were applied. This filtering stage is basically to reduce baseline wandering below 0.5 Hz and high frequency noise above 50 Hz. Amplitude normalization is optional, but helps visually compare signals from different patients.

2.2   Independent Component Analysis

ICA is one of the techniques that solves the blind source separation (BSS) problem [9,10,11]. BSS recovers a set of source signals from the observation of linear mixtures of the sources. These source signals are not directly accessible and have to be extracted or separated from the set of measurable signals or observations. As neither the source signals nor the mixing structure are known, this is referred to as the BSS problem. This problem can be written in a matrix form as

$$X(t) = A \cdot S(t) . \tag{1}$$

$X(t)$ is the vector of acquired signals $X_1(t),...,X_n(t)$ and $S(t)$ is the vector of source signals $S_1(t),...,S_n(t)$. A is called the mixing matrix. The goal of BSS is to estimate $S(t)$ and A from the observations $X(t)$.

The ICA solution to the BSS problem assumes that the sources must be statistically independent and the restriction that the sources must have non-gaussian distributions [11]. The assumption of statistical independence has already been established [6] since during an AF there is an uncoordinated operation of AA and VA [1,12].

2.3   Source Selection for AA

As a result of applying ICA over a 12-lead ECG, a 12-source set is obtained. From this set we have to choose the most representative signal for AA. For this

search, three complementary methods will be applied: kurtosis-based extraction, correlation of the sources to lead V1 of the ECG, and spectral features.

The first method is based on the non-gaussianity of the sources, specially for those with considerable contents of VA. It is not difficult to observe that VA presents a super-gaussian behavior, while the AA behaves as a sub-gaussian random process [6,11]. This non-gaussianity of the sources can be measured by the fourth-order marginal cumulant or kurtosis, that gives a high value, typically above 10 for the VA or super-gaussian distributions and negative values for the AA or sub-gaussian distributions. Thus, the best candidate for AA signal is the source with the lowest kurtosis.

In the second method, a 12 by 12 correlation matrix between ECG leads and sources is generated. In this matrix the column for lead V1 is of particular interest. It is generally accepted that lead V1 captures more atrial activity [12,13]. Thus, the first approach to this method is to correlate lead V1 with all the sources and selects the one with higher correlation index. However, it would be wrong to choose directly the highest value, since the sources with high VA components have high correlation with lead V1, hiding the correlation in the non-QRST segments. To avoid this problem, only the sources with low kurtosis are correlated, obtaining a relatively high correlation in the source that has more similarity with the non-QRST segments of lead V1. Then, the most representative source of AA is chosen when the correlation with lead V1 is greater than the correlation with all the other leads.

However, the fact that lead V1 captures more atrial activity is only a general rule, dependent on both physiological and anatomical variability among patients. That is why it is necessary to have redundant tools for an effective AA extraction. The third method is based in this spectral properties of the fibrillatory waves, which has a distinct peak between 4 and 9 Hz [5,6,14,15]. To indentify this feature, the power spectral density (PSD) is estimated using Welch's averaged modified periodogram method [16]. Then, we have to evaluate a spectral parameter that give us infomation about the relative amount of energy of the spectra in the range between 4 and 9 Hz. This parameter is called spectral concentration (SC) [8] and is defined by

$$SC = \frac{\sum\limits_{f=4}^{9} P(f)}{\sum\limits_{f=0.5}^{50} P(f)} \cdot 100\% \ . \tag{2}$$

$P(f)$ is the PSD of the signal and f the frequency in Hertz. Since we filtered the signals in the band of 0.5 to 50 Hz, the denominator represents the total energy of the signal, while the numerator represents only the energy in the range between 4 and 9 Hz. The criterion applied is that we choose the source with the higher value of SC in the set of 12 sources.

The source chosen by at least two of these three methods is finally selected as the AA signal.

3   Results

With all the 12-lead ECG's signals pre-processed as mentioned in the previous section, FastICA algorithm was applied [17]. For the resulting set of sources the kurtosis method was applied, obtaining in all cases at least 4 sources with kurtosis above 10. This means that those sources have considerable components of VA, represented by a high amplitude QRS complex, as seen in Fig. 1. Table 1 shows the kurtosis found for all sources in all patients studied. We mentioned that AA should have negative kurtosis. For example, in Patient 1, Source 9 is the best candidate for AA signal because it is the source with the lowest kurtosis.

In correlation method, Table 2 shows the selected sources for every patient and the corresponding correlations with the leads. Fig. 2 shows the source selected with this method in Patient 1, with the corresponding segment of lead V1.
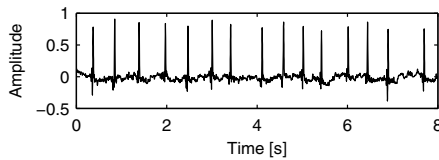
Fig. 1. Source 1 in Patient 1, signal with high kurtosis value. QRS complexes are clearly visible.

**Table 1.** Kurtosis for all the sources in every patient

Source	Patient			
	1	2	3	4
1	24.5156	21.9675	113.8125	21.5941
2	26.3407	23.4926	20.8175	19.9526
3	30.555	19.9116	28.0697	34.0353
4	9.249	10.9939	18.0115	11.7974
5	23.5282	7.9972	7.9839	17.5561
6	1.8372	2.74	15.5995	2.5687
7	1.3795	**-0.6597**	1.7316	1.9107
8	1.3427	1.265	2.0633	1.583
9	**-0.5966**	0.8024	1.2774	**-0.3756**
10	-0.4317	0.4478	0.7189	0.6692
11	0.2128	-0.1481	0.4896	-0.1498
12	-0.0792	0.1545	**0.079**	-0.0366
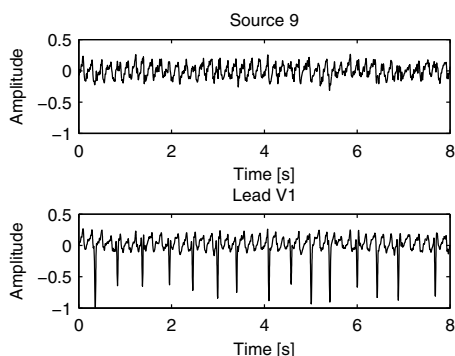
Fig. 2. Comparison between Source 9 and lead V1 in Patient 1. Source 9 has the highest correlation coefficient with this lead.

Table 3 shows the SC parameter for all the sources in the four patients. Fig. 3 shows the PSD of Source 9 (left) in Patient 1, with a SC of 63.2%, and the PSD of Source 1 (right) in the same patient, with a lower SC of only 18.03%. Clearly, Source 9 satisfy the hypothesis mentioned above, with a main peak frequency of 5.86 Hz.

Summarizing, Table 4 shows the results of the three methods for the most representative source of the AA. In this table we see that only in Patient 1 the three methods match. In the other patients one of the methods does not agree. The possible causes for this will be discussed in the next section.
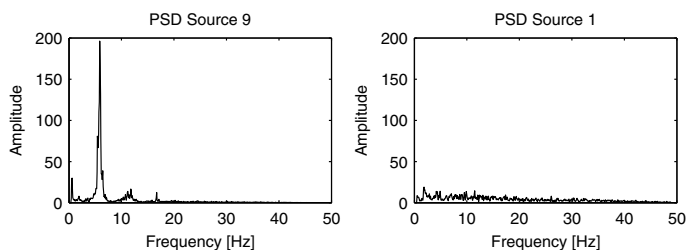
Fig. 3. PSD for Source 9 (left) and Source 1 (right) in Patient 1

4   Discussion

AF is a cardiac alteration that is characterized by uncoordinated atrial activation. On the ECG, AF is seen as the replacement of consistent P waves by rapid oscillations, known as fibrillatory waves, that vary in amplitude and frequency components. Furthermore, anatomical differences between patients give us a source of variabilities in the measurements. This is due to slight variations

**Table 2.** Correlation coefficients for the selected sources

		Patient/Source		
Leads	1/9	2/12	3/12	4/8
I	-0.041	0.247	0.044	-0.100
II	-0.129	0.118	0.066	-0.023
III	-0.125	0.040	0.017	0.038
aVR	0.114	-0.139	-0.054	0.085
aVL	0.076	0.131	0.031	-0.067
aVF	-0.074	0.095	0.082	0.011
V1	**0.439**	**-0.256**	**0.103**	**-0.385**
V2	0.115	0.114	0.002	-0.062
V3	0.067	0.150	0.011	-0.016
V4	0.008	0.174	0.035	-0.001
V5	0.023	0.167	0.026	0.003
V6	0.023	0.141	0.016	0.009

Table 3. Spectral concentration (SC) in percentage for every source in all patients

		Patient		
Source	1	2	3	4
1	18.03	25.45	10.04	14.36
2	16.80	27.76	22.50	23.65
3	16.06	27.35	15.56	14.25
4	10.62	26.32	13.14	14.39
5	11.47	10.13	9.68	21.25
6	12.35	18.03	22.45	9.49
7	57.14	9.86	6.32	32.41
8	18.10	16.46	25.88	26.29
9	**63.20**	13.72	**27.14**	**60.99**
10	19.81	28.86	11.68	16.30
11	24.60	16.13	26.17	11.07
12	14.86	**44.98**	9.65	17.32

Table 4. Summary of results for the three methods. Sources shown are those selected for each method.

|		Method	
Patient	Kurtosis	corr V1	PSD
1	**Source 9**	**Source 9**	**Source 9**
2	Source 7	**Source 12**	**Source 12**
3	**Source 12**	**Source 12**	Source 9
4	**Source 9**	Source 8	**Source 9**
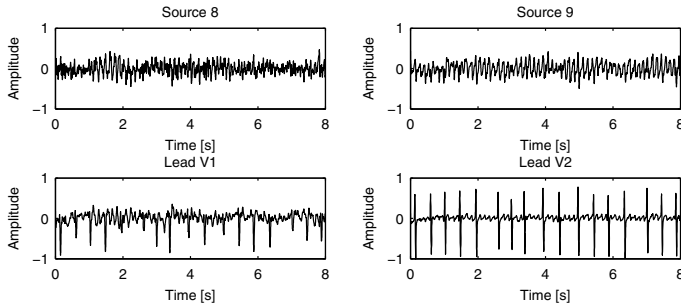
Fig. 4. Comparison between Source 8 and lead V1 (left) and between Source 9 and lead V2 (right). Sources selected by correlation for Patient 4.

in the position of the electrodes in relation to the anatomy of the cardiovascular system of the patient. All these variabilities directly influence the results from the proposed methods, thus we use a consensus between these three selection approaches for AA extraction.

The kurtosis method is an statistical measure, so it is expected that for these fibri-llatory waves we can have variabilities in the estimation of the kurtosis due to the potentially non-stationary nature of the physiological process. For Patient 2 we choose Source 7 due to its lower kurtosis, however, Source 12 (chosen by the other methods) also has a relatively low kurtosis. Although this value is positive, it is relatively close to Source 7 kurtosis, and still far from values of kurtosis from VA. Hence, we can conclude that kurtosis is an effective method to discard the sources that represent the VA, but it is not so accurate to find the source that represents the AA, due to the intrinsic variablities of this activity.

For the correlation method, the hypothesis we use is that lead V1 is the one with more visible AA. As we mentioned, it is possible to have variabilities in anatomical characteristics, resulting in changes of the direction of the electrical vector, projecting lower or higher AA components to the different leads. Lead V1 is not the exception to this problem, so the visibility of AA in this lead will vary for different patients. In Patient 4, Source 8 has the higher correlation with V1, however, the other methods indicate that Source 9 is the most representative source for AA. From the analysis, we found that in this patient, Source 9 has the higher correlation with lead V2. From the above reasoning, we can argue that in this case lead V2 has more visible components of AA. In Figure 4 we show both cases, where clearly lead V2 meets this argument. Also, Source 9 presents a noiseless fibrillatory waves, with a frequency of approximately 7 Hz, matching with the PSD results.

The PSD method is specially sensitive to variabilities in frequency, so it is logical to expect variabilities in the PSD for different cases of AF. However, it is well known that the spectrum of the AA during AF is concentrated between 4 and 9 Hz. In Patient 3 we observe that the SC values for all the sources are under 28%. This means that we have a scattered spectrum, giving lower SC for the range between 4 and 9 Hz. Figure 5 shows the PSD for sources 9 (left)
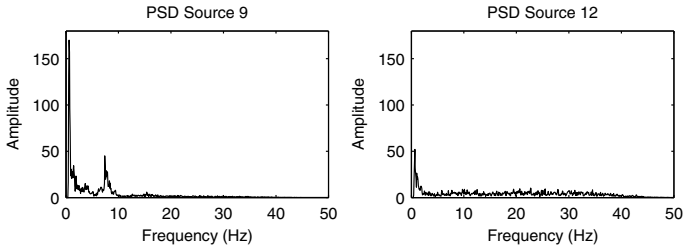
Fig. 5. PSD for sources 9 (left) and 12 (right) in Patient 3. Source 9 have a peak at a frequency of 0.6 Hz, and a secondary peak at 7.4 Hz. Source 12 presents a peak at 0.6 Hz and a scattered distribution in the rest of the spectrum.

and 12 (right) in Patient 3. Source 9 is the one selected by the PSD method, whilst Source 12 is the choice for the kurtosis and correlation methods. It is clear that Source 9 is not the correct signal for AA, since it has a considerable low frequency peak that can not be a component of AA. Aditionally, Source 12 was chosen with a kurtosis value that is not conclusive for the selection, and with correlation method ocurrs the same, where its greatest value is 0.103. Therefore, we have a complicated case for AA extraction, that can easily lead in a complete disagreement of the three methods.

5   Conclusion

In this paper we proposed a reliable extraction of AA during AF. This reliability is based on the use of a consensus between three selection approaches which reduces the effect of the expected variabilities.

The results obtained in this paper are the beginning of a research work which aims to classify AF signals into well-defined subgroups or classes. For the design of a succesful classifier, we need reliable data, from which feature extraction can be done. In that sense, if for a certain AF signal we have a complete disagreement in the results for the three methods proposed, that signal should be discarded.

We trust that the proposed approach will allow for a consistent AA extraction, thus permitting us to move to the next stage of AF classification.

References

1. Fuster, V., Rydén, L.E., Cannom, D.S., et al.: ACC/AHA/ESC 2006 Guidelines for the management of patients with atrial fibrillation. Europace 8(9), 651–745 (2006)
2. Go, A.S., Hylek, E.M., Phillips, K.A., Chang, Y., Henault, L.E., Selby, J.V., Singer, D.E.: Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the anticoagulation and risk factors in atrial fibrillation (atria) study. Journal of the American Medical Association 285(18), 2370–2375 (2001)

3. Stridh, M., Sornmo, L.: Spatiotemporal qrst cancellation techniques for analysis of atrial fibrillation. IEEE Transactions on Biomedical Engineering 48(1), 105–111 (2001)
4. Rieta, J.J., Zarzoso, V., Millet-Roig, J., Garcia-Civera, R., Ruiz-Granell, R.: Atrial activity extraction based on blind source separation as an alternative to qrst cancellation for atrial fibrillation analysis. In: Proc. Computers in Cardiology 2000, pp. 69–72 (2000)
5. Langley, P., Bourke, J.P., Murray, A.: Frequency analysis of atrial fibrillation. In: Proc. Computers in Cardiology 2000. pp. 65–68 (2000)
6. Rieta, J.J., Castells, F., Sanchez, C., Zarzoso, V., Millet, J.: Atrial activity extraction for atrial fibrillation analysis using blind source separation. IEEE Transactions on Biomedical Engineering 51(7), 1176–1186 (2004)
7. Phlypo, R., D'Asseler, Y., Lemahieu, I., Zarzoso, V.: Extraction of the atrial activity from the ecg based on independent component analysis with prior knowledge of the source kurtosis signs. In: Proc. 29th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society EMBS 2007, pp. 6499–6502 (2007)
8. Xiao, J., Chen, Z., Huang, Z.: Atrial activity extraction for atrial fibrillation analysis using mutual information minimization. In: Proc. 1st Int. Conf. Bioinformatics and Biomedical Engineering ICBBE 2007, pp. 916–919 (2007)
9. Cardoso, J.F.: Source separation using higher order moments. In: Proc. Int. Acoustics, Speech, and Signal Processing ICASSP 1989. Conf., pp. 2109–2112 (1989)
10. Comon, P.: Independent component analysis, a new concept? Signal Processing 36, 287–314 (1994)
11. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons, Inc. (2001)
12. Bollmann, A., Husser, D., Mainardi, L., Lombardi, F., Langley, P., Murray, A., Rieta, J.J., Millet, J., Olsson, S.B., Stridh, M., Sörnmo, L.: Analysis of surface electrocardiograms in atrial fibrillation: techniques, research, and clinical applications. Europace 8(11), 911–926 (2006), http://dx.doi.org/10.1093/europace/eul113
13. Thurmann, M., Janney, J.G.: The diagnostic importance of fibrillatory wave size. Circulation 25, 991–994 (1962)
14. Bollmann, A., Kanuru, N.K., McTeague, K.K., Walter, P.F., DeLurgio, D.B., Langberg, J.J.: Frequency analysis of human atrial fibrillation using the surface electrocardiogram and its response to ibutilide. American Journal of Cardiology 81(12), 1439–1445 (1998)
15. Langley, P., Stridh, M., Rieta, J.J., Sornmo, 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. In: Proc. Computers in Cardiology, pp. 29–32 (2002)
16. Welch, P.: The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. IEEE Transactions on Audio and Electroacoustics 15(2), 70–73 (1967)
17. Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. Neural Networks 13(4-5), 411–430 (2000)