Proceedings of the 2005 IEEE
Engineering in Medicine and Biology 27th Annual Conference
Shanghai, China, September 1-4, 2005

QRST cancellation using bayesian estimation for the auricular fibrillation analysis

O Meste[1], N Serfaty[2]
[1]University of Nice-Sophia Antipolis, Lab. I3S, Sophia Antipolis, France
[2]Eurecom institute, Sophia Antipolis, France

Abstract— **The accurate extraction of the atrial fibrillation (AF) signal from the ECG is a challenging task. The performance of its frequency analysis is based on this accuracy. When the atrioventricular node modelling is addressed, the precision of the AF amplitude estimation plays an important role. In this work, we propose an AF estimator based on a bayesian approach. The a priori knowledge necessary to achieve good estimation performances is shown to be weak. The presented method is compared to classical ones such as the Average Beat Substraction (ABS) but also to modern one that is the Blind Source Separation (BSS) technic. The performances of the methods are illustrated on a small set of examples. The conclusion is that the proposed method gives an accurate global estimation of the AF signal but also locally since our results do not exhibit spurious oscillations at the time location of the R wave.**

I. Introduction

When the histogram of the successive R-R intervals exhibits several modes or a large tail departing from the expected gaussian shape it could mean that we are facing an auricular fibrillation (AF) episode [1]. This primary analysis is usually advantageously replaced by a detailed inspection of the V1, V2, V3 leads records in order to improve the diagnosis. In this case, the auricular activity normally represented by a low-frequency gaussian shape when recorded using surface electrode, is replaced by an almost organized cyclic saw-tooth signal. Unlike the normal auricular electrical activity producing the P wave in the ECG, this fibrillation signal is more or less stationary over the whole cardiac cycle. Unfortunately, since its magnitude is low it is superimposed to the ventricular electrical activity represented by the QRST waves.

One concern of the ECG processing during AF episode is the cancellation of the QRST waves in order to analyze accurately the temporal evolution of the spectral content of the AF signal [2]. This analysis is justified by the evident correlation between the spontaneous termination of the episode and the decreasing trend of the AF signal main frequency [3]. Another justification of a precise extraction of the AF electrical activity from the ECG is the modelling of the Atrioventricular junction (AVJ) [1], [4]. It is well known that the success of inducing a ventricular response is related to the characteristics of the electrical impulses which bombard the AVJ. During normal sinus rhythm, a synchronous depolarization wavefront arrives at the AVJ, and hence induce the ventricular response. During atrial fibrillation, loss of spatial coherence in atrial depolarization leads to a reduction of the magnitude of the recorded atrial electrical activity. In this case, the precise measurement of this magnitude will give relevant information to be included in the modelling of the AVJ. In this presented work we will not deal with this modelling but with the accurate evaluation of the atrial electrical activity during paroxysmal AF recorded with surface electrodes.

II. Modelling and processing

In this presented approach, paroxysmal episodes of AF are observed from surface electrodes. The detection of R waves fiducial points from the ECG allows us to define segment with index i defining observation windows with length equal to L including the Q, R, S, T waves. The index j will represent the electrode number. The noise $\mathbf{b}_{ij}$ is constituted by the ambient and contact noise added to the fibrillation signal not synchronous with the R waves. In this model, we assume that in each segment and for each electrode output a linear combination of basic vectors and noise is observed. This hypothesis is commonly used in the QRST cancellation technics [5], and also within the blind source separation (BSS) approach applied to the AF analysis [6]. More explicitly, this involve a matrix constituted by vectors directly attached to the electrical activity produced by the heart, noted $\mathbf{v}_1, \ldots, \mathbf{v}_M$ and some components issued from external variations such as the baseline and offset produced by a not appropriate selection of electrodes. In each segment, the baseline and the offset will be approximated by a linear trend and a constant. Hence, the modelled observation will be given by:

$$\mathbf{x}_{ij} \quad = \quad (\mathbf{v}_1 \ \cdots \ \mathbf{v}_M \ \mathbb{1} \ \mathbf{n})\, \boldsymbol{\theta}_{ij} + \mathbf{b}_{ij} \qquad (1)$$

with $\mathbb{1}$ the unit vector and $\mathbf{n}$ the vector constituted by the values $1 \ldots L$. For each electrode output, vectors $\boldsymbol{\theta}_{ij}$ et $\mathbf{b}_{ij}$ will be modelled as independent gaussian vectors such that $\boldsymbol{\theta}_{ij} \sim \mathcal{N}(\overline{\boldsymbol{\theta}}_j, \mathbf{C}_{\boldsymbol{\theta}_j})$ and $\mathbf{b}_{ij} \sim \mathcal{N}(0, \mathbf{C}_{\mathbf{b}_j})$.

Assuming that the observation windows are perfectly synchronized, the ensemble averaging or synchronous averaging will be computed over the index i, giving:

$$\bar{\mathbf{x}}_j \quad = \quad (\mathbf{v}_1 \ \cdots \ \mathbf{v}_M \ \mathbb{1} \ \mathbf{n})\, \overline{\boldsymbol{\theta}}_j + \boldsymbol{\epsilon}_j \qquad (2)$$

The residual noise $\boldsymbol{\epsilon}_j$ will be neglected for a large number of averaged segments. In addition, the baseline is not synchronized with the fiducial points of the R waves then it appears to be cancelled in the averaging process. This allows us to consider the weighting factor of $\mathbb{1}$ as a function of the offset or recording bias. Hence, after the averaging process, the remaining vectors whose weights are not zero in (2) will be $\mathbf{v}_1, \ldots, \mathbf{v}_M$ and $\mathbb{1}$, defining a reduced dimension matrix $\mathbf{M}_r$. In our application, we have 9 available electrodes from the derivations (V1, ... ,V6, X, Y, Z), producing some redundancy in the measurement of the heart electrical activity. Assuming that the uncertainty $\boldsymbol{\epsilon}_j$ is uncorrelated with other information in $\mathbf{M}_r$, the reduced-rank approach or subspace method leads to a reduction of the dimension of the model (2) when $\mathbf{M}_r$ is not full-rank and in the noisy case. The classical technic [7] consists in using the form (2), which is common to any electrode, to build an averaged matrix $\mathbf{X} = [\bar{\mathbf{x}}_1 \ \ldots \ \bar{\mathbf{x}}_J]$. This matrix is then decomposed using the singular values method (SVD) such that $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$. The least square approximation of $\mathbf{M}_r$ will be given by the vectors in $\mathbf{U}$ corresponding to the highest p singular values in $\mathbf{S}$. This method do not account for the exact knowledge of particular vector forming $\mathbf{M}_r$. This point can be solved as following.

We consider the subspace $\langle\mathbf{K}\rangle$ corresponding to the reduced-rank least square approximation of the subspace $\langle\mathbf{M}_r\rangle$ imposing the partition $\mathbf{K} = [\mathbf{K}_a \ \mathbf{A}]$, with$\mathbf{K}_a$ a full-rank matrix and $\mathbf{A}$ a rank q matrix constituted by known vectors. The least square approximation of the rank p subspace that best describe $\langle\mathbf{X}\rangle$ will correspond to the subspace that span the vectors in $\mathbf{U}$ associated to the $p-q$ highest singular values in $\mathbf{S}$, such that $(\mathbf{I} - \mathbf{A}\mathbf{A}^{\#})\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, appended to the vectors in $\mathbf{A}$. In the text, the symbol $^{\#}$ will stand for the pseudo-inverse.

Proof:

The search of this best least square approximation is equivalent to the minimization of the Frobenius norm $\|\mathbf{X} - [\mathbf{K}_a \ \mathbf{A}]\,\boldsymbol{\theta}\|_F$, that is:

$$\min_{\boldsymbol{\theta}_0,\boldsymbol{\theta}_1,\mathbf{K}_a} \|\mathbf{X} - \mathbf{K}_a\boldsymbol{\theta}_0 - \mathbf{A}\boldsymbol{\theta}_1\|_F \tag{3}$$

The minimization is performed in a first time regards to $\boldsymbol{\theta}_1$ giving $\boldsymbol{\theta}_1 = \mathbf{A}^{\#}(\mathbf{X} - \mathbf{K}_a\boldsymbol{\theta}_0)$. Replacing this solution in (3) gives:

$$\min_{\boldsymbol{\theta}_0,\mathbf{K}_a} \left\| \left(\mathbf{I} - \mathbf{A}\mathbf{A}^{\#}\right)\mathbf{X} - \left(\mathbf{I} - \mathbf{A}\mathbf{A}^{\#}\right)\mathbf{K}_a\boldsymbol{\theta}_0 \right\|_F \tag{4}$$

The classical solution of this minimization consists in the decomposition $(\mathbf{I} - \mathbf{A}\mathbf{A}^{\#})\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ providing the solution $(\mathbf{I} - \mathbf{A}\mathbf{A}^{\#})\mathbf{K}_a = \mathbf{U}\mathbf{S}(p-q)\mathbf{V}^T$. The matrix $(\mathbf{I} - \mathbf{A}\mathbf{A}^{\#})$ being a projector, the matrix to be determined $\mathbf{K}_a$ will be simply constituted by the vectors in $\mathbf{U}$ associated to the highest $p-q$ singular values in $\mathbf{S}$.

In our case $\mathbf{A}$ is the vector $\mathbb{1}$. Then, the reduce model applied to each segment for each sensor is:

$$\mathbf{x}_{ij} = (\mathbf{K}\ \mathbf{n})\,\boldsymbol{\theta}_{ij} + \mathbf{b}_{ij} = \mathbf{H}\boldsymbol{\theta}_{ij} + \mathbf{b}_{ij} \tag{5}$$

In the sequel, the index of segment and sensor will be omitted for the sake of clarity. On the contrary to classical parameters estimation, the noise $\mathbf{b}$ have to be estimated rather than the vector $\boldsymbol{\theta}$. To achieve this task, we propose a bayesian approach based on a conditional probability. In [8], it is shown that the MMSE estimator of a random vector $\mathbf{b}$, conditioned by observations $\mathbf{x}$, is the conditional expectation $E(\mathbf{b}|\mathbf{x})$. Assuming that $\mathbf{b}$ and $\mathbf{x}$ are jointly gaussian, the conditional density $p(\mathbf{b}|\mathbf{x})$ is gaussian and the conditional expectation is :

$$E(\mathbf{b}|\mathbf{x}) = E(\mathbf{b}) + \mathbf{C}_{\mathbf{bx}}\mathbf{C}_{\mathbf{xx}}^{-1}(\mathbf{x} - E(\mathbf{x})) \tag{6}$$

From the initial assumption and using (5) we get:

$$\mathbf{C}_{\mathbf{xx}} = \mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T \text{ and } \mathbf{C}_{\mathbf{bx}} = \mathbf{C}_{\mathbf{b}}, \tag{7}$$

when substituted in (6) gives:

$$\hat{\mathbf{b}}_{MMSE} = E(\mathbf{b}|\mathbf{x}) = \mathbf{C}_{\mathbf{b}}(\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T + \mathbf{C}_{\mathbf{b}})^{-1}(\mathbf{x} - \mathbf{H}\bar{\boldsymbol{\theta}}) \tag{8}$$

Using the matrix inversion lemma, the expression (8) can be rewritten as:

$$\hat{\mathbf{b}}_{MMSE} = (\mathbf{I} - \mathbf{H}(\mathbf{C}_{\boldsymbol{\theta}}^{-1} + \mathbf{H}^T\mathbf{C}_{\mathbf{b}}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}_{\mathbf{b}}^{-1})(\mathbf{x} - \mathbf{H}\bar{\boldsymbol{\theta}}) \tag{9}$$

From this expression it can be deduced that when $\mathbf{C}_{\boldsymbol{\theta}}^{-1} \to 0$ the estimation $\hat{\mathbf{b}}$ tends to:

$$\hat{\mathbf{b}}_{BLUE} = (\mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{C}_{\mathbf{b}}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}_{\mathbf{b}}^{-1})\mathbf{x} \tag{10}$$

This is equivalent to consider that no a priori is used to statistically describe $\boldsymbol{\theta}$ or $\mathbf{C}_{\boldsymbol{\theta}} \to \infty$. Furthermore, The expression (10) of the estimated $\mathbf{b}$ is related to the Best Linear Unbiased Estimator (BLUE) of the parameters vector $\boldsymbol{\theta}$ since it is given by:

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}_{\mathbf{b}}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}_{\mathbf{b}}^{-1}\mathbf{x} \tag{11}$$

Let's recall that it is also a Weighted Least Square estimator assuming that the noise covariance $\mathbf{C}_{\mathbf{b}}$ is known. Thus, by substraction of the estimated $\mathbf{x}$ to the observation, one get the residual noise:

$$\hat{\mathbf{b}} = \mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}} \tag{12}$$

what is equal to (10).

Note that the previous estimators need some specific a priori knowledge on the noise and the parameters. The lonely estimator which do not need such a priori is obtained when replacing $\mathbf{C}_{\mathbf{b}}^{-1}$ by the identity matrix in (11), giving:

$$\hat{\mathbf{b}}_{LS} = (\mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T)\mathbf{x} \tag{13}$$

This estimator, which is an ordinary Least Square (LS) solution, is similar to those proposed in [5] even if the model is not linear regards the parameters. However, in [5] best results are obtained when interpolated noise (or AF signal in our case) is substracted from the observation that also corresponds to use some a priori in the estimation process.

In the blind source separation approach [6], some methods are based on the assumption that sources are orthogonal and others on their independency . In the case of orthogonality the sources ensemble is composed by the atrial fibrillation signal AF ($\mathbf{b}$ in our model) and the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_M$ defined in

(1). If the assumption of orthogonality is verified, vector $\mathbf{b}$ in (5) should be orthogonal to $\mathbf{K}$ since the subspace that $\mathbf{K}_a$ span is the same than those from $\mathbf{v}_1, \ldots, \mathbf{v}_M$. Since $\mathbf{b}$ is free of baseline and offset, $\mathbf{b}$ is then orthogonal to $[\mathbb{1}\ \mathbf{n}]$ which finally establishes the orthogonality between $\mathbf{b}$ and $\mathbf{H}$. This point should explain that the estimated $\mathbf{b}$ in (13) should be identical to the true $\mathbf{b}$ up to a noise effect. If the orthogonality is not verified it will produce a filtered $\mathbf{b}$ with a time variant filter corresponding to $\mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$.

The knowledge of the covariance matrix $\mathbf{C}_{\boldsymbol{\theta}}$ and $\mathbf{C}_{\mathbf{b}}$ seems to be crucial for the MMSE (8) and only $\mathbf{C}_{\mathbf{b}}$ is needed for the BLUE (10). The previous assumption related to orthogonality is clearly stated with real signals and the model in (5) could not fit to any segment. Thus the proposed approach will based on a two steps estimation. When considering only one sensor records without loss of generality, it is summarized as:

- For all records, compute the LS estimated noise (or AF) $\mathbf{b}_{LS}$ in (13) taken as entries for the estimated covariance matrix $\hat{\mathbf{C}}_{\mathbf{b}}$
- For all records, compute the BLUE estimated noise $\hat{\mathbf{b}}_{BLUE}$ in (10) substituting $\mathbf{C}_{\mathbf{b}}$ by $\hat{\mathbf{C}}_{\mathbf{b}}$

Note that this procedure can be iterated since the second step should improve the estimation of the $\mathbf{b}$ and consequently improve the accuracy of the estimated $\mathbf{C}_{\mathbf{b}}$.

The covariance $\hat{\mathbf{C}}_{\mathbf{b}}$ will be formed accounting for the symmetric Toeplitz structure of covariance matrix whose elements will be the average of each individual autocorrelation function of estimated noise $\mathbf{b}_{LS}$ (13). This procedure assume that each individual autocorrelation function is represented by the average one and that the aim of averaging is to reduce some noise (not the AF) effect. In other words, the use of the averaged autocorrelation function or the deduced covariance matrix is justified if the AF signal is almost regular from one segment to an other.

When the AF signal is no longer stationary, a recursive procedure can be used to update the averaged autocorrelation function $\overline{ACF}_i(m)$ of the segment i, as following:

$$\overline{ACF}_i(m) = (1-\lambda)\overline{ACF}_{i-1}(m) \quad (14)$$

$$+\lambda \frac{1}{N} \sum_{n=0}^{N-m-1} \hat{b}_{LS,i}(n+m)\hat{b}_{LS,i}(n) \quad (15)$$

The adaptation coefficient λ is chosen in the range [0,1] and the initial averaged autocorrelation function $\overline{ACF}_1(m)$ is computed with the single $\hat{\mathbf{b}}_{LS,1}$. In the following real application, this adaptation will not be taken into account because of the almost stationary characteristics of these particular AF signals. However, the good results obtained under this assumption show the robustness of the method.

III. Results with a real database

In order to illustrate the performances of the methods, we present representative examples taken from two patients with chronic atrial fibrillation. Results from $\hat{\mathbf{b}}_{MMSE}$ are not significantly different from $\hat{\mathbf{b}}_{BLUE}$ when the unknown matrix and vector $\mathbf{C}_{\boldsymbol{\theta}}$ and $\overline{\boldsymbol{\theta}}$ are estimated using (11). In each figure, the extracted AF activity is shown in one selected segment of duration equal to 550 ms and for the V1 lead. Each R-wave fiducial point corresponds to the time index 100 of each segment and is estimated from the Y derivation. Figure 1 and figure 3 represent two segments including a zoomed portion of the derivation V1, including the QRST waves and the AF signal, the AF estimated using $\hat{\mathbf{b}}_{BLUE}$ (10) and $\hat{\mathbf{b}}_{LS}$ (13), respectively for the two patients. Figures 2 and 4 represent the same information where the well known ABS results [9] have been substituted to the V1 derivation. In figure 5, the $\hat{\mathbf{b}}_{BLUE}$ result is compared to the best source estimated using two blind source separation (BSS) approaches [6]. Note that for BSS methods the ICA and PCA results have been magnified and vertically shifted to fit the $\hat{\mathbf{b}}_{BLUE}$ result for visual comparison. Comparing results from $\hat{\mathbf{b}}_{BLUE}$, $\hat{\mathbf{b}}_{LS}$ and ABS, it is clear that $\hat{\mathbf{b}}_{BLUE}$ exhibits much less spurious oscillations, around index 100, than other estimators. The periodic pattern of the AF is also very well preserved using $\hat{\mathbf{b}}_{BLUE}$, on the contrary to others. As expected in figure 5, the extracted sources (AF) using BSS approach are much more noisy and also exhibit artifacts around index 100.

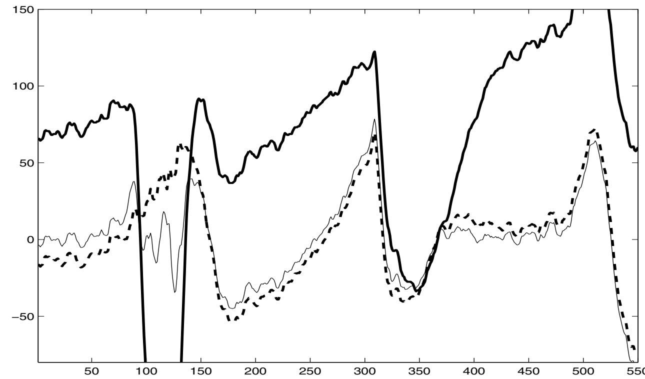

Fig. 1. Segment I from the V1 lead for the first patient: A zoomed portion of V1 lead (thick line), the estimated AF signals ($\hat{\mathbf{b}}_{BLUE}$: dashed line, $\hat{\mathbf{b}}_{LS}$: thin solid line)

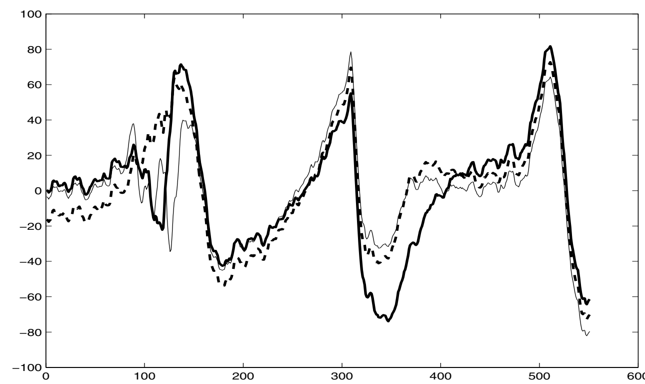

Fig. 2. Segment I from the V1 lead for the first patient: the estimated AF signals ($\hat{\mathbf{b}}_{BLUE}$: dashed line, $\hat{\mathbf{b}}_{LS}$: thin solid line, ABS result: thick line)

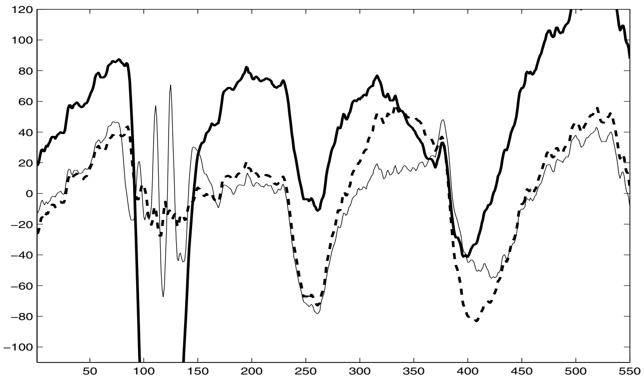

Fig. 3. Segment II from the V1 lead for the second patient: A zoomed portion of V1 lead (thick line), the estimated AF signals ($\hat{\mathbf{b}}_{BLUE}$: dashed line, $\hat{\mathbf{b}}_{LS}$: thin solid line)

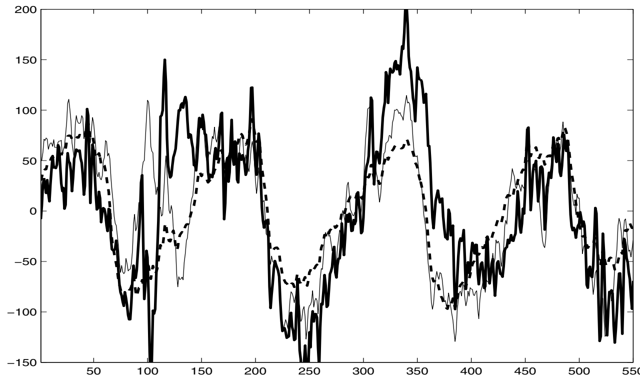

Fig. 5. Segment III for the second patient: the estimated AF signals ($\hat{\mathbf{b}}_{BLUE}$ for the V1 lead: dashed line, the best source BSS-PCA: thick line, the best source BSS-ICA: thin solid line)

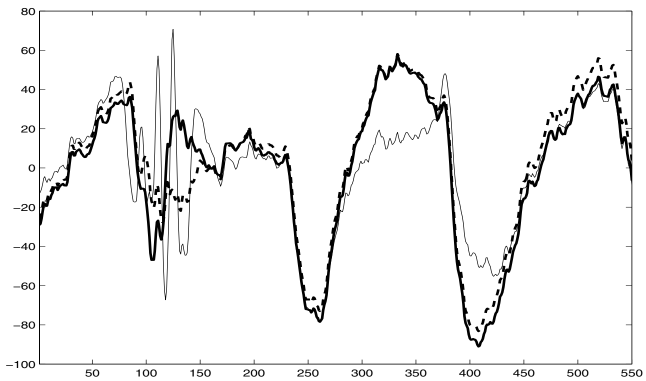

Fig. 4. Segment II from the V1 lead for the second patient: the estimated AF signals ($\hat{\mathbf{b}}_{BLUE}$: dashed line, $\hat{\mathbf{b}}_{LS}$: thin solid line, ABS result: thick line)

IV. Conclusion

A new QRST cancellation technic has been proposed in this work. The first step of the global approach is to reduce the redundancy of information contained in the modelling matrix using approximation. The effect of such approximation is a rank reduction taking into account some knowledge on the matrix. This first step, which is similar to a principal component analysis, allows a reduction of the parameters number leading to decrease the variance of their estimation. Starting from a Bayesian approach, we have obtained a particular estimator assuming that the a priori knowledge is only related to the AF statistics. On a small set of data we have shown that our AF estimator is accurate and robust regards to classical methods and modern ones. These promising results will allow some future extension of the atrioventricular node modelling where the electrical wavefront produced by the atrium is recorded using surface electrodes.

Acknowledgement

The authors would like to express their gratitude to Leif Sörnmo for providing the real data.

References

[1] R. J. Cohen, R. D. Berger, and T. E. Dushane, "A quantitative model for the ventricular response during atrial fibrillation," *IEEE Trans Biomed Eng*, vol. 30, no. 12, pp. 769–781, 1983.

[2] Q. Xi, A. V. Sahakian, and S. Swiryn, "The effect of qrs cancellation on atrial fibrillatory wave signal characteristics in the surface electrocardiogram," *Journal of Electrocardiology*, vol. 36, no. 3, pp. 243–249, 2003.

[3] S. Petrutiu1, A.V. Sahakian, J. Ng, and S. Swiryn, "Analysis of the surface electrocardiogram to predict termination of atrial fibrillation: The 2004 computers in cardiology/physionet challenge," *Proc. Computers in Cardiology*, vol. 31, pp. 105–108, 2004.

[4] J. Lian, D. Mussig, and V. Lang, "Computer modelling of ventricular rhythm during atrial fibrillation and ventricular pacing," *Proc. Computers in Cardiology*, vol. 31, pp. 29–32, 2004.

[5] M. Stridh and L. Sörnmo, "Spatiotemporal qrst cancellation techniques for analysis of atrial fibrillation," *IEEE Trans Biomed Eng*, vol. 48, no. 1, pp. 105–111, 2001.

[6] J. J. Rieta, F. Castells, C. Sanchez, and V. Zarzoso J. Millet, "Atrial activity extraction for atrial fibrillation analysis using blind source separation," *IEEE Trans Biomed Eng*, vol. 51, no. 7, pp. 1176–1186, 2004.

[7] L. L. Scharf, "The svd and reduced rank signal processing," *Signal Processing*, vol. 25, pp. 113–133, 1991.

[8] S. Kay, *Fundamentals of statistical signal processing, estimation theory*, Prentice-Hall Inc., 1993.

[9] J. Slocum, E. Byrom, L. McCarthy, A. Sahakian, and S. Swiryn, "Computer detection of atrioventricular dissociation from the surface electrocardiogram during wide qrs tachycardia," *Circulation*, vol. 72, pp. 1028–1036, 1985.