33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Detection and Removal of Ventricular Ectopic Beats in Atrial Fibrillation Recordings Via Principal Component Analysis

Arturo Martínez*, Raúl Alcaraz and José J. Rieta, *Member, IEEE*

Abstract— **Ectopic beats are early heart beats with remarkable large amplitude that provoke serious disturbances in the analysis of electrocardiograms (ECG). These beats are very common in atrial fibrillation (AF) and are the source of important residua when the QRST is intended to be removed. Given that QRST cancellation is a binding step in the appropriate analysis of atrial activity (AA) in AF, a method for ventricular ectopic beats cancellation is proposed as a previous step to the application of any QRST removal technique. First, the method discriminates between normal and ectopic beats with an accuracy higher than 99% through QRS morphological characterization. Next, the most similar ectopic beats to the one under cancellation are clustered and serve to get their eigenvector matrix by principal component analysis. Finally, the highest variance eigenvector is used as cancellation template. The reduction ectopic rate (RER) has been defined to evaluate the method's performance by using templates generated with 5, 10, 20, 40 or 80 ectopics. Optimal results were reached with the 5 most similar complexes, yielding a RER higher than 5.5. In addition, a decreasing RER trend was noticed as the number of considered ectopics for cancellation increased. As conclusion, given that ectopics presented a remarkable variability in their morphology, the proposed cancellation approach is a robust ectopic remover and can notably facilitate the later application of any QRST cancellation technique to extract the AA in the best conditions.**

I. Introduction

THE appearance of ventricular ectopic beats in the electrocardiogram (ECG) is a sign of disturbance in the depolarization process, disorganizing the blood pumping function of the ventricles and preceding, in many cases, malignant cardiac arrhythmias [1]. Although this sign of decreased heart function is helpful in the prediction of life-threatening arrhythmias, such as ventricular fibrillation or tachycardia [2], its presence makes difficult the study of other cardiac diseases from the ECG. In this respect, the presence of ectopic beats perturbs the impulse pattern initiated by the sinoatrial node and implies that RR intervals, adjacent to an ectopic beat, cannot be used for heart rate variability (HRV) analysis [3], which is a tool commonly used in the evaluation of the autonomic nervous system [4]. Similarly, ventricular ectopic beats are a frequent and challenging problem in atrial fibrillation (AF) long-time monitoring that still is unsolved.

AF is the supra-ventricular arrhythmia most commonly diagnosed in clinical practice and affects up to 1% of the general population and up to 15% of the population over 80 years [5]. The atrial activity (AA) signal extraction under the best conditions is crucial in the study of the electrophysiological processes that underlie AF, such as refractory periods, autonomic response, drug effects, etc. [6]. However, this extraction requires nonlinear signal processing techniques since the atrial and ventricular activity overlap spectrally and, therefore, cannot be separated by linear filtering [6]. Average beat subtraction (ABS) is the most widespread technique for AA extraction, especially for long-time Holter recordings, which routinely present no more than two or three leads. This method relies on the assumption that the average beat can represent approximately each individual beat [6]. However, QRST morphology is often subject to minor changes caused by respiration, patient movement, etc., and, therefore, QRST residua are often present in the estimated AA [7]. In the same way, the presence of ventricular ectopics, which exhibit a very different morphology to the normal beats, provokes important residua in the AA signal. To overcome this problem, some variants of ABS generate one cancellation template for each particular beat morphology present in the ECG [8]. Other ABS modifications have generalized this approach by creating one template for each beat, showing that optimal cancellation of normal complexes is achieved when the most similar 20 or 30 beats to those under cancellation are used to generate its template [9]. However, none of these techniques have paid a special attention dealing to ectopic beats, which are originated by different mechanisms and present different morphological characteristics than normal beats. Additionally, no study has been developed yet about the ideal number of ectopics allowing to provide a template for optimized cancellation.

In this work, a method for ventricular ectopic beats cancellation in single-lead AF signals is proposed as a previous step to the AA extraction with ABS or any of its variants. A cancellation template for each individual ectopic is generated by exploiting, through principal component analysis (PCA), the mutual information available in the set of its most similar beats. Given that this method requires discrimination between normal and ectopic beats, an algorithm based on QRS morphological characterization is also proposed. Finally, a study about the optimal number of ectopics to generate the most accurate cancellation template is presented.

II. Methods

A. Database and preprocessing

To validate the proposed detector and canceller of ectopic beats, twenty 10 hours-length ECG segments with AF and a considerably high density of ectopic beats were extracted

*A. Martínez and R. Alcaraz are with the Innovation in Bioengeeniering Research Group, University of Castilla-La Mancha, Campus Universitario, 16071, Cuenca, Spain (email: arturo.martinez@uclm.es, raul.alcaraz@uclm.es).

J. J. Rieta is with the Biomedical Synergy, Electronic Engineering Department, Universidad Politécnica de Valencia, Spain (email: jjrieta@ieee.org).

from 24-h Holter recordings corresponding to 15 and 5 patients with paroxysmal and persistent AF, respectively. Although three leads (II, aVF and V1) were recorded, only V1 was considered in the study. The recording was upsampled to 1 kHz in order to improve the time alignment accuracy between each ectopic and its cancellation template [4]. Additionally, this lead was filtered to remove baseline wander, high frequency noise and powerline interference.

B. Detection of ectopic beats

A morphological characterization of each ventricular beat was developed for the detection of ectopic beats. Thus, QRS complexes were firstly detected by using a method based on the Phasor Transform (PT) [10]. This transform was used to convert each instantaneous sample of the preprocessed ECG ($y(n)$) into a phasor with magnitude $M(n)$ and phase $\varphi(n)$, such that $PT\{y(n)\} = M(n)e^{j\varphi(n)}$. Given that maximum instantaneous phase variation in the ECG was found for the QRS complexes, these waves were located as the segments exceeding a threshold. The R-peak of the ith beat (r_i) was marked as the maximum magnitude point within the corresponding segment [10]. Thereafter, two points around the R-peak, $\gamma_{i_{QRS-}}$ and $\gamma_{i_{QRS+}}$, were established. They were defined as the closer points to the R-peak in which amplitude was 30% of this peak amplitude, i.e. $y(\gamma_{i_{QRS-}}) = y(\gamma_{i_{QRS+}}) = 0.3 \cdot y(r_i)$, see Fig. 1. Next, a forward 15 ms window relative to the R-peak was established to search a local minimum, which was referred as r'_i. Around this point, $\rho_{i_{QRS-}}$ and $\rho_{i_{QRS+}}$ were marked following the same approach used for the $\gamma_{i_{QRS-}}$ and $\gamma_{i_{QRS+}}$ establishment, see Fig. 1. Finally, five descriptors to characterize the QRS morphology of the beat were defined as:

- Area of the QRS complex (Ar_i), i.e., the sum of the preprocessed ECG samples between the identified boundary points, i.e.: $Ar_i = \sum_{k=\gamma_{i_{QRS-}}}^{\gamma_{i_{QRS+}}} y(k)$.
- Number of samples between $\gamma_{i_{QRS+}}$ and $\gamma_{i_{QRS-}}$ (D_i), i.e.: $D_i = \gamma_{i_{QRS+}} - \gamma_{i_{QRS-}}$.
- Number of samples between $\rho_{i_{QRS+}}$ and $\rho_{i_{QRS-}}$ (D'_i), i.e.: $D'_i = \rho_{i_{QRS+}} - \rho_{i_{QRS-}}$.
- Amplitude of the R-peak, i.e. $y(r_i)$.
- Absolute value of the local minimum amplitude following the R-peak, i.e. $|y(r'_i)|$.

To classify a ventricular beat as normal or ectopic, its descriptors were compared with those corresponding to the average of its preceding five beats detected as normals. In this way, slight progressive morphological variations in normal beats were considered. A decision scheme based on a sequence of five rules was built such that, a beat was classified as normal when none of the rules were satisfied, see Fig. 2. In contrast, a beat was considered as ectopic when any of the rules was fulfilled. These rules were experimentally obtained and can be mathematically expressed as inequalities:

- Rule 1: $Ar_i \geq 1.8 \cdot \overline{Ar}$,
- Rule 2: $D_i \geq 2 \cdot \overline{D}$,
- Rule 3: $D'_i \geq 2 \cdot \overline{D'}$,
- Rule 4: $y(r_i) \geq 2.5 \cdot \overline{y(r)}$,

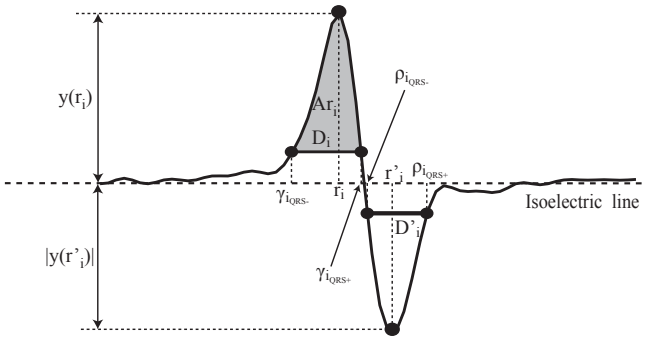

Fig. 1. Graphical representation of the defined descriptors to characterize morphologically the ith QRS complex.

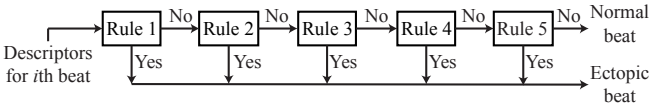

Fig. 2. Decision scheme to classify beats as normal or ectopic based on a sequence of rules.

- Rule 5: $|y(r'_i)| \geq 2.5 \cdot |y(r')|$,

$\overline{Ar}$, $\overline{D}$, $\overline{D'}$, $\overline{y(r)}$ and $\overline{|y(r')|}$ being the descriptors averaged for the five normal beats preceding those under classification, i.e., the ith beat. Obviously, this approach requires, as a starting point, manual selection of five normal beats, in a similar way to previously published techniques for heartbeat classification [11]. After several experimental tests, this number of beats was obtained as a trade-off between classification accuracy and computational burden.

C. Cancellation of ectopic beats

Graphical inspection of the ECG allows us to appreciate that ventricular ectopic beats usually exhibit recurrent patterns, although different QRST morphologies as well as variations in their waveforms may occur. Therefore, ventricular ectopic beats show a high degree of temporal redundancy which could be exploited for their cancellation making use of PCA. In order to take advantage of this ergodicity, the Q-wave onset and the T-wave offset for each ectopic were localized making use of a PT-based delineator [10]. To delimit all the ectopic beats with the same duration, the median values of the QR ($\overline{QR}$) and RT ($\overline{RT}$) intervals were obtained and the Q-wave onset and the T-wave offset for the ith ectopic were defined as $q_{i-} = r_i - \overline{QR}$ and $t_{i+} = r_i + \overline{RT}$, respectively. Hence, this beat was represented by the column vector

$$\begin{aligned}\mathbf{x}_i &= \left[\begin{array}{cccc} y(q_{i-}) & y(q_{i-}+1) & \ldots & y(t_{i+}-1) \quad y(t_{i+}) \end{array}\right]^{\mathsf{T}} \\ &= \left[\begin{array}{cccc} \mathbf{x}_i(1) & \mathbf{x}_i(2) & \ldots & \mathbf{x}_i(L-1) \quad \mathbf{x}_i(L) \end{array}\right]^{\mathsf{T}}, \end{aligned} \quad (1)$$

where L is the number of samples.

In order to generate the cancellation template of an ectopic beat, only the N most similar complexes to it were chosen. Similarity among ectopics was obtained in terms of the cross-correlation index (κ), which can be defined for the ith and

jth beats as

$$\kappa = \frac{E[\mathbf{x}_i \widetilde{\mathbf{x}}_j]}{\sigma_i \sigma_j}, \tag{2}$$

where $E[\cdot]$ is the expectation operator and σ_i and σ_j are the standard deviations of both beats. It is noteworthy that the jth ectopic was adapted in amplitude to the ith beat for a better comparison of morphologies, such that

$$\widetilde{\mathbf{x}}_j = \frac{y(r_i) - y(r'_i)}{y(r_j) - y(r'_j)} \mathbf{x}_j. \tag{3}$$

Temporal redundancy of this set of N complexes was exploited by applying PCA. Thus, for the ith beat, these N complexes were assembled in a matrix $\mathbf{X}_i \in \Re^{L \times N}$, such that

$$\mathbf{X}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iN}]. \tag{4}$$

Note that all the beats were temporally aligned using their R-peak timings. The principal components of this matrix associated with PCA were obtained by singular value decomposition (SVD) [12], such that

$$\mathbf{X}_i = \mathbf{U}_i \mathbf{S}_i \mathbf{V}_i^{\mathsf{T}}, \tag{5}$$

where $\mathbf{U}_i \in \Re^{L \times N}$ is a unitary matrix so that $\mathbf{U}_i \mathbf{U}_i^{\mathsf{T}} = \mathbf{I}$, being $\mathbf{I}$ the identity matrix, $\mathbf{S}_i \in \Re^{N \times N}$ is a diagonal matrix, and $\mathbf{V}_i \in \Re^{N \times N}$ fulfills $\mathbf{V}_i \mathbf{V}_i^{\mathsf{T}} = \mathbf{I}$. The matrix $\mathbf{U}_i = [\mathbf{u}_{i1}, \ldots, \mathbf{u}_{iN}]$ contains the N normalized principal components of $\mathbf{X}_i$, their cross-correlation coefficients being null. The eigenvalues, i.e. the amplitude coefficients corresponding to the principal components, corresponds with the diagonal elements of $\mathbf{S}_i$. Given that the first principal component $\mathbf{u}_{i1}$, i.e., the highest variance eigenvector, could be considered as the mother representation of the ith beat [9], [13], this vector was taken as ventricular cancellation template. Thus, the subtraction of this ectopic was obtained as

$$\widehat{\mathbf{y}}_i = \mathbf{x}_i - \widetilde{\mathbf{u}}_{i1}, \tag{6}$$

$\widetilde{\mathbf{u}}_{i1}$ being the vector $\mathbf{u}_{i1}$ adapted in amplitude to the ith beat in the way expressed in (3). Once all the ventricular ectopic beats were cancelled out, AA could be extracted from the signal $\widehat{y}(n)$ by using any of the QRST cancellation techniques published to date.

D. Performance assessment

The performance of the proposed ectopics detector was compared with two previously published techniques, such as Krasteva & Jekova's [11] and Acar et al.'s [14] methods. Details on their implementation can be obtained in the corresponding literature. Given that the used database did not contain annotations, recordings were manually annotated by expert cardiologists for distinguishing between normal and ectopic beats. These annotations were used to evaluate the reliability of the analyzed detectors. Thus, sensitivity (S_e) and positive predictivity (P^+) were computed as

$$S_e = \frac{TP}{TP + FN}, \tag{7}$$

$$P^+ = \frac{TP}{TP + FP}, \tag{8}$$

TABLE I

PERFORMANCE OF THE ANALYZED ECTOPIC BEATS DETECTORS

Method	TP	FP	FN	$S_e(\%)$	$P^+(\%)$
Proposed in this work	31848	696	446	98.62	97.86
Krasteva & Jekova [11]	30327	29714	1967	93.91	50.51
Acar et al. [14]	10153	8303	22141	31.44	55.01

TP being the number of true positives (i.e., ectopic beats correctly classified), FN the number of false negatives (i.e., ectopic beats incorrectly classified) and FP the number of false positives (i.e. normal beats incorrectly classified).

On the other hand, when real ECG recordings are analyzed, the only evidence for a successful cancellation of any ectopic beat is the absence of its residua [9]. Unfortunately, there is no parameter in the literature able to robustly quantify the existence of QRST residua in the resulting signal after cancellation. Thereby, a new parameter to estimate the reduction rate of an ectopic beat after its cancellation is next proposed. Thus, the reduction ectopic rate (RER) of the ith beat was defined as

$$RER_i = \sqrt{\frac{\sum_{k=q_i-}^{t_i+} y(k)^2}{\sum_{k=q_i-}^{t_i+} \widehat{y}(k)^2}}. \tag{9}$$

Fig. 3 shows visually how this parameter behaves. As can be appreciated, the higher the ectopic beat residua, the lower the RER values.

III. RESULTS

The ectopic beats detection in terms of sensitivity and positive predictivity obtained by the proposed algorithm, the Krasteva & Jekova's method [11] and the Acar et al.'s technique [14] are shown in Table I. A notably better performance can be observed for the proposed algorithm than for the two other methods. Thus, the proposed algorithm classified correctly 99.88% of the analyzed beats, whereas for the other methods the classification rate was significantly more reduced. It should be noted that a total number of 997387 beats were analyzed, of which 32294 were identified as ectopics by expert cardiologists.

On the other hand, each ectopic was cancelled out with a template generated from its 5, 10, 20, 40 and 80 more similar beats in order to obtain a first idea about the optimal N value for the most accurate cancellation. The RER, expressed as mean $\pm$ standard deviation (std), for each tested N value is presented in Table II. As can be seen, the best cancellation was obtained when the most 5 similar complexes were selected, achieving a reduction of the ectopic beats by a factor slightly higher than 5.5, in average. In addition, a decreasing trend in the RER was observed when N increased. As a graphical summary, Fig. 3 shows a ECG segment after ectopic beats cancellation with $N = 5$, 20 and 80. In agreement with the obtained RER values, the highest residua can be observed in the residual signal after cancellation with $N = 80$.

TABLE II

RER FOR ECTOPIC BEATS CANCELLATION WITH DIFFERENT N VALUES

N	5	10	20	40	80
mean	5.55	5.23	4.77	3.88	2.95
std	1.21	1.06	0.92	0.95	0.96

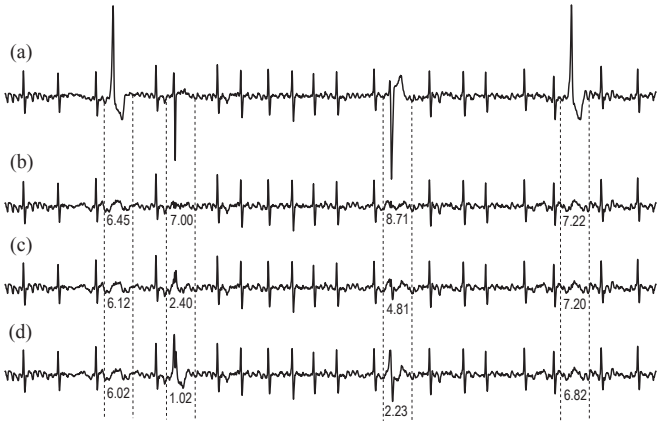

Fig. 3. Original ECG (a) and residual signal after ectopic cancellation considering the most (b) 5, (c) 40 and (d) 80 similar complexes to those under cancellation to generate the template. It can be seen that the RER, indicated numerically, can reliably reflect the reduction of ectopic beats after cancellation.

IV. DISCUSSION AND CONCLUSIONS

The proposed method requires the previous identification of ectopic beats. Although different techniques for discerning between normal and ventricular ectopic beats can be found in the literature, such as Krasteva & Jekova's [11] and Acar et al.'s [14] methods, they have not been specifically designed for AF recordings. Indeed, for the classification of beats, these two algorithms use descriptors based on the regularity of RR and PQRST intervals, which are inappropriate for AF signals, given their high ventricular irregularity [7]. In addition, they were validated with MIT-BIH Arrhythmia Database, in which only 9% of the recorded time corresponds with AF episodes [15], and a subjects group without AF, respectively. This fact could justify the high accuracy, greater than 98%, shown for both methods from the databases with which were originally validated and the poor sensitivity and positive predictivity provided in this work. Overall, a robust algorithm with a high ability of discriminating between normal and ventricular ectopic beats, based only on QRS morphological characterization, has been proposed and validated with AF recordings.

A previous study on the selection of the normal beats allowing to generate the most accurate QRST cancellation template has shown that the highest quality AA extraction is obtained when the most similar complexes to those under cancellation are selected [9]. Thereby, in this work, for each ectopic beat only its most N similar beats were considered for the generation of the template. In disagreement with the optimal number of normal beats found to generate the most accurate QRST cancellation template, i.e. between 20 and 30 [9], the highest reduction rate of ectopic beats was reached by using the 5 most similar complexes, the RER being approximately of 5.5. Additionally, higher residua of ectopic beats were noticed in the residual signal when a greater number of complexes was considered for cancellation. As a consequence, although more variability in morphologies of ectopic than normal beats is appreciated, a successful cancellation of ectopic beats is obtained by considering, as ventricular template, the highest variance eigenvector associated with the set of the most similar ectopics to those under cancellation. Nonetheless, it should be remarked that this work only presents preliminary results, which will be completed with more detailed analyses.

V. ACKNOWLEDGMENTS

This work was supported by the projects TEC2010–20633 from the Spanish Ministry of Science and Innovation and PPII11–0194–8121 and PII1C09–0036–3237 from Junta de Comunidades de Castilla La Mancha.

REFERENCES

[1] G. A. Ng, "Treating patients with ventricular ectopic beats," *Heart*, vol. 92, no. 11, pp. 1707–12, Nov 2006.

[2] H. J. J. Wellens, A. P. Gorgels, and H. de Munter, "Cardiac arrest outside of a hospital: how can we improve results of resuscitation?" *Circulation*, vol. 107, no. 15, pp. 1948–50, Apr 2003.

[3] K. Solem, P. Laguna, and L. Sörnmo, "An efficient method for handling ectopic beats using the heart timing signal," *IEEE Trans Biomed Eng*, vol. 53, no. 1, pp. 13–20, Jan 2006.

[4] A. Bollmann, D. Husser, L. Mainardi, F. Lombardi, P. Langley, A. Murray, J. J. Rieta, J. Millet, S. B. Olsson, M. Stridh, and L. Sörnmo, "Analysis of surface electrocardiograms in atrial fibrillation: techniques, research, and clinical applications," *Europace*, vol. 8, no. 11, pp. 911–26, Nov 2006.

[5] W. B. Kannel, P. A. Wolf, E. J. Benjamin, and D. Levy, "Prevalence, incidence, prognosis, and predisposing conditions for atrial fibrillation: population-based estimates." *Am J Cardiol*, vol. 82, no. 8A, pp. 2N–9N, Oct 1998.

[6] L. Sörnmo, M. Stridh, D. Husser, A. Bollmann, and S. B. Olsson, "Analysis of atrial fibrillation: from electrocardiogram signal processing to clinical management," *Philos Transact A Math Phys Eng Sci*, vol. 367, no. 1887, pp. 235–53, Jan 2009.

[7] S. Petrutiu, J. Ng, G. M. Nijm, H. Al-Angari, S. Swiryn, and A. V. Sahakian, "Atrial fibrillation and waveform characterization. a time domain perspective in the surface ecg," *IEEE Eng Med Biol Mag*, vol. 25, no. 6, pp. 24–30, 2006.

[8] M. Lemay, J.-M. Vesin, A. van Oosterom, V. Jacquemet, and L. Kappenberger, "Cancellation of ventricular activity in the ecg: evaluation of novel and existing methods," *IEEE Trans Biomed Eng*, vol. 54, no. 3, pp. 542–6, Mar 2007.

[9] R. Alcaraz and J. J. Rieta, "Adaptive singular value cancelation of ventricular activity in single-lead atrial fibrillation electrocardiograms," *Physiol Meas*, vol. 29, no. 12, pp. 1351–69, Dec 2008.

[10] A. Martínez, R. Alcaraz, and J. J. Rieta, "Application of the phasor transform for automatic delineation of single-lead ecg fiducial points," *Physiol Meas*, vol. 31, no. 11, pp. 1467–85, Nov 2010.

[11] V. Krasteva and I. Jekova, "Qrs template matching for recognition of ventricular ectopic beats," *Ann Biomed Eng*, vol. 35, no. 12, pp. 2065–76, Dec 2007.

[12] G. H. Golub and C. V. van Loan, *Matrix Computations*, 2nd ed. Baltimore, Md, USA: The Johns Hopkins University Press, 1989.

[13] F. Castells, P. Laguna, L. Sörnmo, A. Bollmann, and J. Millet, "Principal component analysis in ecg signal processing," *EURASIP J. Adv. Signal Process.*, vol. 2007, no. 1-22, 2007.

[14] B. Acar, I. Savelieva, H. Hemingway, and M. Malik, "Automatic ectopic beat elimination in short-term heart rate variability measurement," *Comput Methods Programs Biomed*, vol. 63, no. 2, pp. 123–31, Oct 2000.

[15] G. B. Moody and R. G. Mark, "The MIT-BIH arrhythmia database on CD-ROM and software for use with it," *Computers in Cardiology*, vol. 17, pp. 185–188, 1990.