AUTOMATIC IDENTIFICATION AND CORRECTION OF ECTOPIC BEATS IN SHORT HEART PERIOD TIME SERIES

T. Bassani*, D.M. Simpson**, and A. Beda*

* Department of Electronic Engineering, Federal University of Minas Gerais, Belo Horizonte, Brazil

** Institute of Sound and Vibration Research, University of Southampton, Southampton, UK

email: ale.beda@gmail.com

Abstract: A novel automatic method identifying and correcting ectopic beats (EB) in heart period time series (RR) is proposed. The approach is based on the improvement in predicting RR as autoregressive process when progressively removing beats supposed to be EB (EBi). Reliability of the approach was tested assessing RR series recorded in 19 healthy subjects at rest, before and after substituting RR samples with EB in varying percentage. The method properly recognised the absence of EB and the percentage of values identified as EBi resulted corresponding to percentage of inserted EB. EBi were found correctly matching EB samples with percentages larger than 80%. Insertion of EB into RR affected power spectral indices increasing absolute power at high frequency (0.15-0.5Hz). This effect was cancelled by deleting EBi from RR. The present method avoids the need of user intervention and is accountable as a valid option in cases of high time consuming RR analysis.

Keywords: Heart rate variability, artefacts, automatic ectopic beat correction.

Introduction

Heart period, usually computed as the time interval between two consecutive R peaks on the ECG (RR), is routinely analysed to derive spectral indices to indirectly assess short-term cardiovascular regulation [1].

The presence of ectopic beats (EB) into RR is known as potentially interfering with RR power spectrum analysis [2,3] thus requiring adequate processing. When assessing power spectral indices of RR, deletion of EB was found to perform as well as or better than more complex methods to correct EB with appropriate values [4,5].

Removing EB via manual editing of RR can be very time consuming, especially in the clinical context, when many patients are considered or when long ECG recordings are available (i.e. Holter recordings). In these situations, automatic methods of EB rejection can be advantageous.

With this aim, it becomes of primary importance to correctly identify EB automatically. Many authors have proposed approaches to identify EB in RR [6-8]. The main limitation of applying these existing methods for automatic identification is related to the need of user intervention in setting suitable indices or decisional thresholds for EB recognition.

The present work presents a novel approach in recognising EB based on the autoregressive prediction of RR and in particular on the significant improvement in predicting RR when progressively removing beats supposed to be EB. This approach is completely automatic and does not require the user to select the values of a-priori indices or thresholds.

The reliability of the proposed approach was tested by adding artificial EB in RR series recorded from healthy subjects at rest. The performance of the method was quantified by the percentage of EB correctly identified and by the error in the spectral indices of RR resulting from not removing the undetected EB.

Materials and methods

Experimental protocol – Nineteen (19) healthy subjects (mean age 25 ± 26.5 [SD], 11 women) were evaluated while sitting comfortably and relaxed without talking. Every subject was monitored for 5 minutes. The protocol adhered to the principles of the Declaration of Helsinki and was approved by the local ethical committee [9].

Data acquisition and processing – ECG was recorded continuously with a sampling frequency of 1000 Hz. The RR time series was obtained from the ECG using an automated algorithm followed by manual editing in order to avoid artefacts [9]. RR={RR(i), i=1,..., N} was extracted on a beat-to-beat basis, where N is the series length. Stationary sequences of 250–300 consecutive RR values (i.e. recordings of few minutes) were analysed.

Addition of artificial ectopic beats – Randomly choosing in RR a pair of consecutive RR(j) and RR(j+1) samples, artificial ectopic beat and consequent compensatory beat were added replacing RR(j) with $0.25\cdot(RR(j)+RR(j+1))$ and RR(j+1) with

0.75·(RR(j)+RR(j+1)) respectively. Basically, inserting one ectopic beat implies to replace a couple of RR samples preserving RR mean unaltered.

The random substitution of 5%, 10%, 15% and 20% of samples of the original RR with EB was generated for every subject, resulting in the modified series RR# where #=5,10,15,20.

Automatic correction of ectopic beats – Given *y*=RR or *y*=RR#, *y* series was modelled as a linear time invariant autoregressive process whose parameters were estimated via least-squares method [10]. The ordinary least-squares problem can be described in matrix notation as:

$$y = Y\beta + w \quad (1)$$

$$\hat{\beta} = (Y^T Y)^{-1} Y^T y \quad (2)$$

$$y - Y\hat{\beta} = e \quad (3)$$

where $y=\{y(i), i=1,..., N\}$ is an Nx1 vector, Y is an Nxp matrix containing in the i-th row the p past values of $y(i)$, β is a px1 vector of parameters to be estimated, w is an Nx1 error term with zero mean and Gaussian distribution. According to Eq. (3) prediction error e is an Nx1 vector representing the difference between y and its prediction based on $\hat{\beta}$ estimated in Eq. (2). The best model order p, ranging from 5 to 15, was optimized according to Akaike criterion [11]. The ability of the model in fitting y is measured according to the mean squared prediction error (MSPE) calculated as variance of e. The lower is MSPE the best is model fitting.

In order to identify the presence of EB in y, values of e most distant from its mean value were considered as candidate (i.e. those beats that the model could not predict well). Accordingly, tail values of distribution of e were considered and the pool of values equal in number to 5% of N, the half of which in each tail, were labelled as EB. Rows of y and Y corresponding to the identified EB values were then removed from Eq. (2) and $\hat{\beta}$ was anew estimated thus recalculating e and MSPE. The procedure was iteratively repeated removing up to 30% of N samples with step of 5%, resulting in progressively decreasing values of MSPE. In this context, the value of MSPE(k) (k=1,..,7 being the number of iteration, with 1 the initial prediction step) is a function of the percentage of removed EB.

With the aim of assessing the significance of MSPE(k) reduction between consecutive iterations, confidence intervals were estimated using Monte Carlo simulations. Considering e calculated at initial step, a set of 100 Gaussian distributed series with same mean and variance of e was created, the resulting MSPE calculated. The 2.5th and 97.5th percentiles of the distribution of the simulated MSPE were taken as upper and lower confidence limits respectively. The process was repeated at each iteration.

MSPE(k) decreases at every iteration due to the progressive cancellation of tail values. In order to recognize the iteration beyond which the improvement in diminishing MSPE(k) becomes negligible, the maximum variation of first derivative of MSPE(k) was calculated ($MSPE_S$). If $MSPE_S$ was found falling above the lower threshold it was then considered as significant improvement in predicting y when removing the corresponding values.

Values removed in correspondence to $MSPE_S$ were classified as EB (EBi). Since *y*=RR# the corrected series $RR\#^C$ was then obtained by deleting in RR# the EBi.

Indices of comparison – In order to assess the reliability of the proposed method we computed: i) percentage of correction (PC_S) - the ratio between the number of samples identified as EBi and the length of RR series evaluated; ii) correction matching (CM_S) - percentage of EBi that actually match "true" EB.

In order to assess the effect of adding EB on RR series and of deleting EBi, the power spectral density (PSD) of RR# and $RR\#^C$ was estimated in the LF (0.05–0.15Hz) and HF frequency band (0.15–0.5 Hz) and compared with PSD of original RR. PSD was estimated using a parametric approach, using an autoregressive model of RR series estimated via Levinson–Durbin recursion [12]. The model order ranged between 5 and 15 and was optimized according to the Akaike criterion [11]. The whiteness of the residuals was checked.

The LF and HF absolute powers of RR (RR_{LF} and RR_{HF}), RR# ($RR\#_{LF}$ and $RR\#_{HF}$) and $RR\#^C$ ($RR\#^C_{LF}$ and $RR\#^C_{HF}$) were obtained as the total power of PSD in the respective frequency band [13].

Statistical analysis – Significant differences between $RR\#_{LF}$ and $RR\#^C_{LF}$ and between $RR\#_{HF}$ and $RR\#^C_{HF}$ were evaluated using 2-way repeated measures ANOVA or Friedman's rank sum test if normality condition of data was not satisfied. In case of significant differences, comparison with RR_{LF} and RR_{HF} in the specific band was performed with paired t-test or Wilcoxon rank-sum test whereas necessary. Results were considered significant if p-value<0.05.

Results

As reported in Table 1 the percentage of values recognised as EBi (PC_S) was found to be <1% when evaluating original RR series and highly similar with percentage of inserted EB when assessing RR#'s. The percentage of EBi that correctly matched with inserted EB exhibited values larger than 80% in relation to all RR#'s (see CM_S in Table 1).

Table 1: mean values of percentage of correction (PC_S) and of EBi correction matching (CM_S) calculated in correspondence to RR and RR#'s.

	RR	RR5	RR10	RR15	RR20
PC_S	<1%	5%	9%	15%	17%
CM_S	-	83%	90%	92%	82%

As depicted in Fig. 1 absolute powers of RR# and $RR\#^C$ were not significantly different at LF (see Fig.1a) and showed values similar to RR_{LF}. When assessed at HF (see Fig.1b) absolute powers of RR# and $RR\#^{C}$ where significantly different and $RR\#_{HF}$'s were always significantly larger than RR_{HF} (see Fig.1b) with the difference increasing with the number of EB introduced. On the contrary $RR\#^{C}_{HF}$'s were similar to RR_{HF} with the solely exception of $RR20^{C}_{HF}$ (see Fig.1b).

Discussion

The results show that the proposed method is able to handle the EB. Indeed, the mean value of PC_S was found <1% when assessing the original RR's (Table 1). High matching between PC_S and RR#'s are inferable from Table 1. This result points out the correctness of the relation between percentage of RR# values identified as EBi and the percentage of RR values substituted with EB to create RR#'s. This supporting result is further confirmed assessing the percentage of matching between EBi and EB effectively inserted obtaining RR#'s (see CM_S in Table 1). On this regard CM_S exhibited mean values always larger than 80%.

An example of processing a RR series and subsequent RR10 is presented in Fig. 2. When evaluating RR (Fig.2d) the progressive decrease of MSPE(k) remains bounded inside its confidence interval (Fig.2e) thus appropriately avoiding identification of false EBi. On the contrary when processing RR10 (Fig.2a) significant improvement in the prediction is correctly observed in correspondence to 3rd iteration (see $MSPE_S$ in Fig.2b) and thus when 10% of values of RR10 (i..e. EBi) have been removed. Matching between EBi and EB that were artificially added obtaining RR10 is well recognizable checking dot markers in Fig.2a.

When assessing PSD indices the substitution of RR values with EB did not seem to affect LF band (see Fig.1a). The solely exception of $RR20_{LF}$ was not sufficient in determining significant differences. Both $RR\#_{LF}$ and $RR\#^{C}_{LF}$ exhibited values similar to RR_{LF} (see

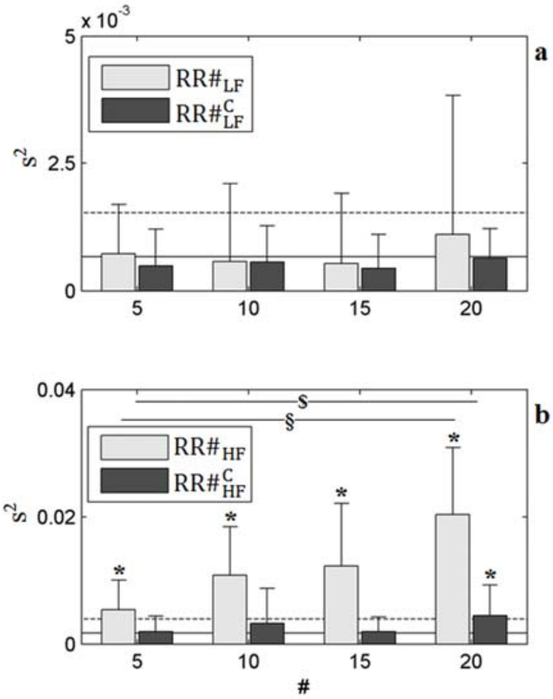

Figure 1: Absolute powers at LF (a) and HF (b) of RR# (light grey) and $RR\#^{C}$ (dark grey) expressed as mean+SD. RR_{LF} and RR_{HF} are depicted as mean (solid line) plus SD (dashed line) in (a) and (b) respectively. \$, indicates significant difference between $RR\#_{HF}$ and $RR\#^{C}_{HF}$ with p<0.001. §, indicates significant difference among $RR\#_{HF}$'s repetitions with p<0.001. *, indicates significant differences of $RR\#_{HF}$'s and $RR\#^{C}_{HF}$'s in comparison to RR_{HF} at p<0.05.

Fig.1a). This result suggests that deleting EB from RR series has a negligible effect in LF power estimation.

A completely different outcome was observed evaluating absolute power at HF (see Fig.1b). The significant increasing in $RR\#_{HF}$ is clearly correlated

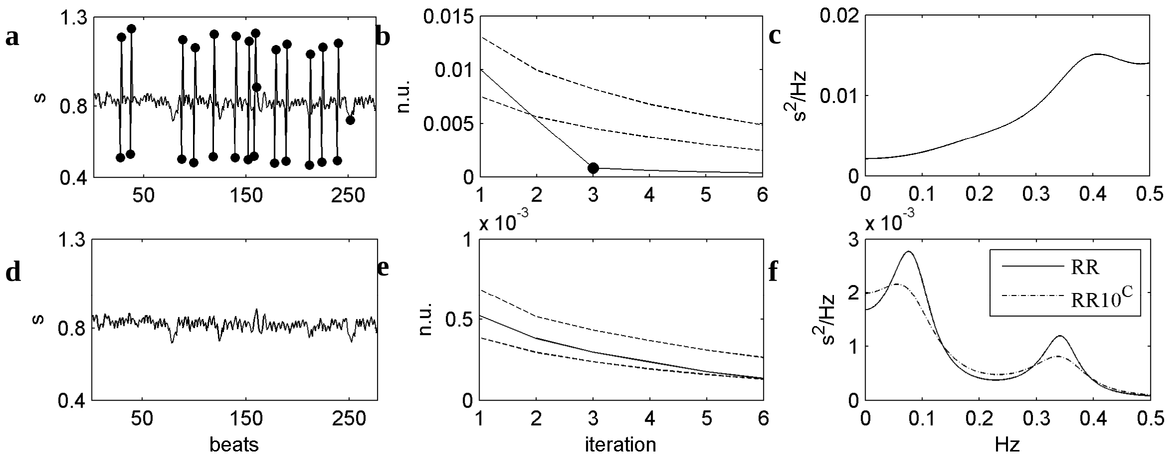

Figure 2: In the upper line example of RR10 (a) with EB found in correspondence to $MSPE_S$ indicated as dot markers; MSPE(k) of RR10 (b) depicted as solid line with confidence interval as dashed lines and $MSPE_S$ as dot marker; RR10 PSD (c). In the lower line original RR (d) used to create corresponding RR10; MSPE(k) of RR (e); PSD of RR and of $RR10^{C}$ (f).

with the increase in the percentage of EB inserted into RR. This result was somehow expected since EB represent spurious high frequency oscillations. An example of this effect is shown in Fig.2: introducing 10% of EB in RR (Fig.2a) results in a considerable increase in HF power (Fig.2c) if compared with PSD of original RR series (Fig.2f).

When assessing absolute power of corrected series $RR\#^{C}$ in the HF band (see $RR\#^{C}_{HF}$ in Fig.1b) values exhibited by $RR\#^{C}_{HF}$ were found significantly different with respect to $RR\#_{HF}$. In addition all $RR\#_{HF}$'s were found significantly larger than RR_{HF} (see Fig.1b). This result seems to indicate that the automatic deletion of EBi can rectify the estimation of RR absolute HF power in presence of EB. This finding is exemplified in Fig.2f, showing that the PSD of corrected series $RR10^{C}$ and of original RR are quite similar .

It is worth noting that the 25%-75% ratio between premature contraction and compensatory pause chosen to generate ectopic beats was arbitrarily based on the observation of several ectopic ECG tracings. Tests using a wider range of ratios are needed to fully validate the method.

Conclusion

The proposed method proved to be suitable in identifying the presence of EB in short time RR and can represent a reliable approach when automatically processing RR series correction. The main advantage of the present approach consists in avoiding the need of user intervention and seems thus considerable as a valid option in cases of high time consuming RR analysis. Further developments concern the comparison of automatic identification with manual editing in presence of real EB and the response when dealing with long time RR series.

References

[1] Cohen M.A. and Taylor J.A., Short-term cardiovascular oscillations in man: measuring and modelling the physiologies. J. Physio, 542, 669–683, 2002.

[2] Berntson G.G. and Stowell J.R., ECG artifacts and heart period variability: don't miss a beat! Psychophysiology, 35(1):127-32, 1998.

[3] Wen F. and He F.T., An efficient method of addressing ectopic beats: new insight into data preprocessing of heart rate variability analysis. J Zhejiang Univ Sci B, 12(12): 976–982, 2011.

[4] Lippman N., Stein K.M., Lerman B.B., Comparison of methods for removal of ectopy in measurement of heart rate variability. AJP - Heart and Circulatory Physiology, 267, H411-H418, 1994.

[5] Storck N., Ericson M., Lindblad L., Jensen-Urstad M. Automatic computerized analysis of heart rate variability with digital filtering of ectopic beats. Clin Physiol, 21(1):15-24, 2001.

[6] Berntson G.G., Quigley K.S., Jang J. and Boysen S.T., An approach to artifact identification: Application to heart period data. Psychophysiology, 27, 586–598, 1990.

[7] Mateo J., Laguna P., Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal. IEEE Trans. Biomed. Eng., 50(3):334-343, 2003.

[8] Liu C.Y., Li L.P., Zhao L.N., Zheng D.C., Li P., Liu C.C., A Combination Method of Improved Impulse Rejection Filter and Template Matching for Identification of Anomalous Intervals in RR Sequences. Journal of Medical and Biological Engineering, 32(4): 245-250, 2012.

[9] Beda A., Jandre F.C., Phillips D.I., Giannella-Neto A. and Simpson D.M., Heart-rate and blood-pressure variability during psychophysiological tasks involving speech: Influence of respiration. Psychophysiology, 44, 767–778, 2007.

[10] Soderstrom T. and Stoica P., System identification, Prentice Hall, Englewood Cliffs, NJ, 1988.

[11] Akaike H., A new look at the statistical novel identification. IEEE Transactions on Automatic Control, 19: 716-23, 1974.

[12] Kay S.M. and Marple S.L., Spectrum analysis: a modern perspective. Proceedings of the IEEE, 69: 1380-418, 1981.

[13] Baselli G., Porta A., Rimoldi O., Pagani M., and Cerutti S., Spectral decomposition in multichannel recordings based on multivariate parametric identification. IEEE Trans Biomed Eng, 44: 1092-101, 1997.