Accurate R Peak Detection and Advanced Preprocessing of Normal ECG for Heart Rate Variability Analysis

Devy Widjaja[1], Steven Vandeput[1], Joachim Taelman[1], Marijke AKA Braeken[2], Renée A Otte[2], Bea RH Van den Bergh[2], Sabine Van Huffel [1]

[1]Department of Electrical Engineering, ESAT-SCD, Katholieke Universiteit Leuven, Belgium
[2]Department of Developmental Psychology, Universiteit van Tilburg, The Netherlands

Abstract

Heart rate variability (HRV) analysis is well-known to give information about the autonomic heart rate modulation mechanism. In order to avoid erroneous conclusions, it is of great importance that only sinus rhythms are present in the tachogram. Therefore, preprocessing of the RR interval time series is necessary. This paper presents an advanced automated algorithm to preprocess RR intervals obtained from a normal ECG.

Validation of this algorithm was performed on one hour ECG signals of 20 pregnant women. R peaks before and after preprocessing were manually revised for spurious and missed R peak detections. Before preprocessing, more than 1% of the detected R peaks were incorrect while preprocessing corrected more than 94% of these errors leading to an overall error rate of 0.06%. Our automated preprocessing technique therefore restricts the manual data check to the absolute minimum and allows a reliable HRV analysis.

1. Introduction

Heart rate variability (HRV) analysis is well-known to give information about the autonomic heart rate modulation mechanism. In order to avoid erroneous conclusions, it is of great importance that only sinus rhythms are present in the tachogram. Therefore, preprocessing of the RR interval time series is necessary [1, 2]. R peaks have to be detected accurately in the ECG and missed peaks or false peaks have to be corrected. Also ectopic or supraventricular beats have to be removed.

A commonly used automated preprocessing technique is the 20% filter [3]. RR intervals differing more than 20% of the previous interval are replaced by the average value of the 5 preceding and 5 following intervals: $RR_{new} = \frac{1}{10}\sum_{i=-5,i\neq 0}^{i=5} RR_{old+i}$. However, this simple technique is not accurate. Moreover, Figure 1 shows that this preprocessing technique would introduce errors in the RR interval time series as the difference between 891 ms and 683 ms is 30%.

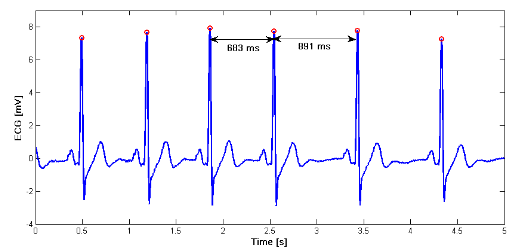

Figure 1. Correct R peak detections. Preprocessing according to the 20% filter would introduce errors.

Instead of manual revision of all the detected R peaks, a new automated method is proposed which restricts the manual data check to the absolute minimum and allows a reliable HRV analysis.

2. Methods

2.1. Data acquisition

The data for this study are part of a larger project that investigates the influence of stress and anxiety during pregnancy. For this project 140 women, aged 18-40, are recruited from 10 to 12 weeks gestation onwards. Inclusion criteria are: no current substance abuse problems, no severe psychiatric problems and no pregnancy-associated medical problems such as diabetes or hypertension. The participants are subjected to Holter monitoring, during which the ECG is recorded at 1000 Hz by the Vrije Universiteit - Ambulatory Monitory System [4]. One hour of the Holter recordings of 20 women were randomly selected for this study. During the selected recordings, the women were awake and were doing their normal daily activities.

Table 1. Performance measures (# R peaks)

Before preprocessing	After preprocessing
# correctly detected (TP_1)	# correctly removed (TN_2)
# falsely detected (FP_1)	# falsely removed (FN_2)
# falsely undetected (FN_1)	# falsely unremoved (FP_{2a})
	# correctly added (TP_2)
	# falsely added (FP_{2b})

2.2. Preprocessing algorithm

During automated R peak detection, which is performed by the Pan-Tompkins algorithm [5], false detections occur mostly due to noise. Often, noise causes spurious R peak detections, which is harmless, since no information is lost. On the other hand, undetected R peaks always result in the loss of information [1].

Figure 2 shows the flowchart of the preprocessing algorithm. The proposed technique attempts to recover correct RR intervals by summing consecutive small intervals and thus removing spurious R peaks. To check whether an interval is too small, a reference RR interval (RR_{ref}), which is empirically set as a weighted average of three previous RR intervals, is used for comparison. In case of a small RR interval ($RR_i < 0.7RR_{ref}$), a summation with preceding and following RR intervals is optimized in the way that the resulting sum is closest to RR_{ref}. If no optimal summation can be found ($RR_{sum} < 0.7RR_{ref}$ or $1.3RR_{ref} < RR_{sum}$), the small RR interval is flagged for manual revision. Therefore, these flagged intervals are rejected as part of a new reference interval. Too large RR intervals ($RR_i > 1.8RR_{ref}$) are evenly divided in smaller intervals in order to obtain RR intervals that are closest to RR_{ref}.

2.3. Performance measures

To quantify the performance of the preprocessing algorithm, a comparison of the detected R peaks before and after preprocessing is made with respect to the manually annotated R peaks. Table 1 shows the performance measures, which categorize the detected and undetected R peaks before and after preprocessing. Using these measures, the performance of the preprocessing algorithm is quantified by means of the error rate, which is defined as the ratio between the number of errors and the actual number of R peaks (TP_1+FN_1). Before preprocessing, the error rate is expressed as:

$$error_{before} = \frac{FP_1 + FN_1}{TP_1 + FN_1}.$$

After preprocessing, the error rate is determined by:

$$error_{after} = \frac{FN_2 + FP_{2a} + FP_{2b} + FN_1 - TP_2}{TP_1 + FN_1}.$$

3. Results and discussion

Figure 3 demonstrates the performance of the preprocessing algorithm. Instead of altering the value of small RR intervals, like the 20% filter does, this preprocessing technique succeeds in recovering the correct RR intervals.

The results of the performance measures on the 1h ECG recordings of the 20 pregnant women are shown in Table 2. Application of the Pan-Tompkins algorithm resulted in an error rate ($error_{before}$) of 1.0936%. This indicates that the Pan-Tompkins algorithm detected almost 99% of the R peaks correctly. Preprocessing corrected more than 94% of these errors leading to an overall error rate ($error_{after}$) of 0.0624%. Remark that the flagged intervals are not yet manually corrected. However, in 12 out of the 20 cases, there were no errors detected after preprocessing.

Table 2. Results of the performance measures (# R peaks)

Before preprocessing	After preprocessing
TP_1 = 108987	TN_2 = 1151
FP_1 = 1185	FN_2 = 29
FN_1 = 7	FP_{2a} = 34
	TP_2 = 3
	FP_{2b} = 1

Table 3 demonstrates the importance of preprocessing by displaying the mean and standard deviation of the RR intervals before and after preprocessing as well as the actual values. Data of only 5 women are presented, including the data of the 3 women who scored the worst after preprocessing (ID 503, 508 and 509). In those cases, *meanNN* and *SDNN* before preprocessing show large deviations of the actual values. Consequently, the use of these values may lead to erroneous conclusions. However, preprocessing corrected most of the errors, leading to very small deviations, even in the worst cases.

Table 3. Mean and standard deviation of the RR intervals: the actual values and the values before and after preprocessing

	meanNN [ms]			SDNN [ms]		
ID	*actual*	*before*	*after*	*actual*	*before*	*after*
503	657.68	622.97	657.68	89.41	153.59	89.54
504	687.64	687.38	687.64	84.31	85.22	84.31
507	772.03	759.97	771.86	71.02	107.40	71.18
508	694.18	660.19	694.18	59.18	136.29	59.55
509	683.89	653.95	683.89	37.49	120.17	37.65

These results indicate the excellent performance of the preprocessing technique. Manual revision of the flagged RR intervals is not even strictly necessary. This is especially favorable during Holter monitoring.

De Chazal *et al.* proposed a preprocessing algorithm that was also based on summing small RR intervals and evenly dividing large RR intervals, but no detailed algorithm was presented [6]. Their preprocessing algorithm

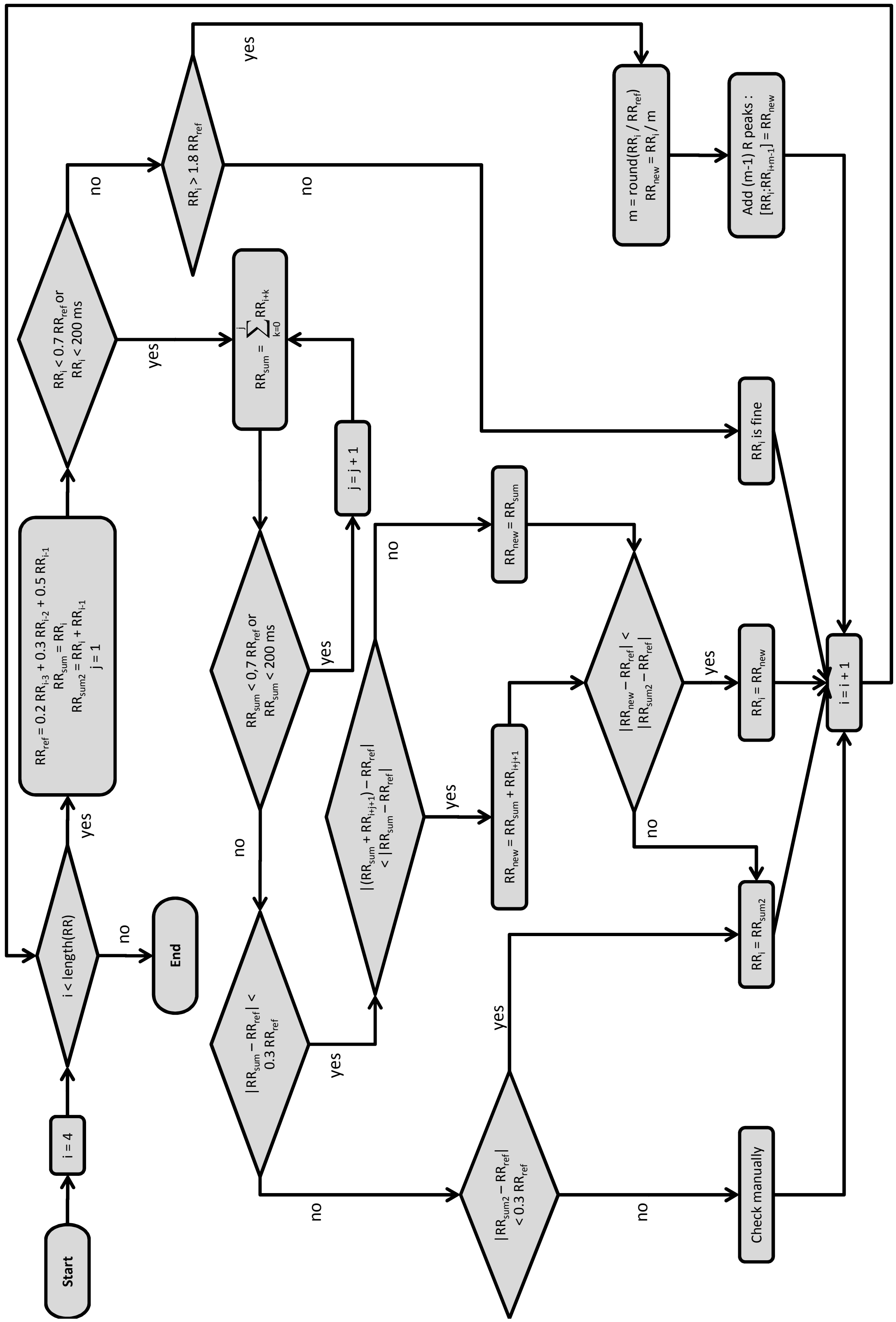

Figure 2. Flowchart of the preprocessing algorithm.

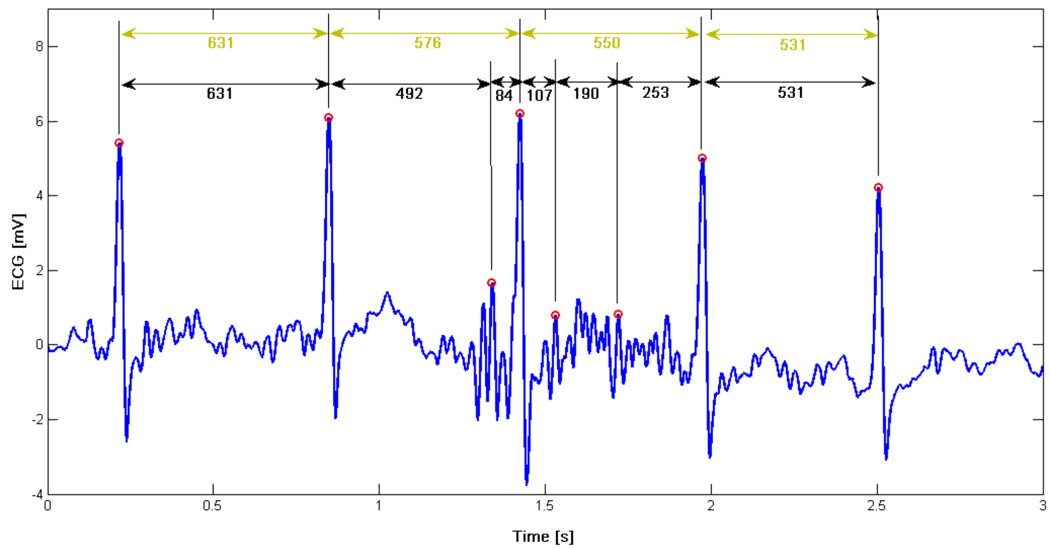

Figure 3. Performance of the proposed preprocessing algorithm (*black:* RR intervals before preprocessing [ms], *green:* RR intervals after preprocessing [ms]).

resulted in only 98.6% correct R peak detections. It is however difficult to compare these results because a different validation set was used.

4. Conclusions

This paper presented a new algorithm for automated preprocessing. This algorithm managed to recover correct RR intervals by using information on previous RR intervals intelligently, resulting in 99.94% correct RR intervals after preprocessing. Also, the mean and standard deviation of preprocessed RR intervals showed minimal deviations of the actual values, restricting the manual data check to the absolute minimum and allowing a reliable HRV analysis.

Acknowledgements

Research supported by
- Research Council KUL: GOA Ambiorics, GOA MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), IDO 05/010 EEG-fMRI, IDO 08/013 Autism, IOF-KP06/11 FunCopt, several PhD/postdoc & fellow grants;
- Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, 'Dynamical systems, control and optimization', 2007-2011); ESA PRODEX No 90348 (sleep homeostasis)
- EU: FAST (FP6-MC-RTN-035801), Neuromath (COST-BM0601)

References

[1] Berntson G, Bigger Jr J, Eckberg D, Grossman P, Kaufmann P, Malik M, Nagaraja H, Porges S, Saul J, Stone P, et al. Heart rate variability: origins, methods, and interpretive caveats. Psychophysiology 1997;34(6):623–648.

[2] Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. European Heart Journal 1996;17(3):354–381.

[3] Kleiger R, Miller J, Bigger Jr J, Moss A. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. The American Journal of Cardiology 1987;59(4):256–262.

[4] Vrijkotte T, van Doornen L, de Geus E. Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. Hypertension 2000;35(4):880.

[5] Pan J, Tompkins W. A real-time QRS detection algorithm. IEEE Transactions on Biomedical Engineering 1985;230–236.

[6] De Chazal P, Heneghan C, Sheridan E, Reilly R, Nolan P, O'Malley M. Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. IEEE Transactions on Biomedical Engineering 2003; 50(6):686–696.

Address for correspondence:

Devy Widjaja
K.U.Leuven, ESAT/SISTA
Kasteelpark Arenberg 10
B-3001 Leuven-Heverlee
Belgium
devy.widjaja@esat.kuleuven.be