Reliability Loss with Sampling Rate Reduction

Paulo Sousa[1 2], Rute Almeida[1 2 3], Marta João Silva[2 4 5], Ana Paula Rocha[1 2]

[1] Faculdade de Ciências, Universidade do Porto, Porto, Portugal
[2] Centro de Matemática da Universidade do Porto (CMUP), Porto, Portugal
[3] BSICoS Group, I3A, Universidad de Zaragoza & CIBER-BBN, Zaragoza, Spain
[4] Unidade de Cuidados Intensivos Pediátricos, Centro Hospitalar São João, Porto, Portugal
[5] Faculdade de Medicina, Universidade do Porto, Porto, Portugal

Abstract

High sampling frequency is not usually available in hospital monitoring systems, what can limit the usefulness of the data, namely for repolarization measures.

In this work the reliability of beat-to-beat measures using low sampling frequency is quantified with respect to the original high sampling rate. ECG recordings originally at 1000 Hz, including intensive care patients data, were downsampled to 500 Hz and to 250 Hz. Automatic delineation was applied to extract beat-to-beat intervals. The reliability was measured considering concordance correlation coefficient (CCC) and intraclass correlation coefficient (ICC); the Information–based measure of disagreement (IBMD) was used for agreement quantification.

High reliability and low disagreement were generally found. Using the conventional interpretation excellent consistency was found at 250 HZ for more than 68%, 50%, 50% and 58% of the cases for RR, RT, QT and QT_c, respectively. Results indicate that at 250 Hz RR, RT, QT and QT_c measures are still reliable compared with the values at 1000 Hz.

1. Introduction

The autonomic nervous system (ANS) plays an important role in the human response to various internal and external stimuli, and its dysfunction may complicate the clinical course of critically ill (CI) patients. The golden standard of ANS characterization is heart rate variability (HRV) quantification assessed from measures over the electrocardiogram (ECG). Decreased HRV has been related with disease severity and outcome in a broad range of critical illnesses [1]. Dysfunctions of the ANS in CI patients with Acute Brain Injury (ABI) lead to changes in HRV accessed from 1000 Hz ECG recordings, which appear to be particularly marked in patients subsequently declared in Brain Death (BD) [2]. Uncoupling of QT/RR also measured at 1000 Hz has showed clinical potential in differentiating the progression of the disease and the outcome within children with ABI [3].

High sampling frequency is not usually considered in hospital monitoring systems, what can limit the usefulness of the data, namely for repolarization measures. In this work clinically useful time measures derived from ECG recordings at different sampling frequencies are compared, trying to evaluate their consistency and agreement.

The degree of association between two sets of data is usually evaluated by the correlation coefficient. However, a correlation coefficient only relies in how two sets of scores vary together (i.e. proportionally), not the extent of agreement between them, therefore their values can be misleading [4]. On the other hand, using the same scale as the measurements, measures of agreement (A) quantify how close two measurements made on the same subject are [5]. The reliability (R) reports the overall consistency of a measure and evaluates the precision of the measurements in terms of the ability to distinguish between subjects or groups [4, 6]. Nevertheless, the currently used reliability indexes, Lin's concordance correlation coefficient (CCC) and intraclass correlation coefficient (ICC), are also affected by any disagreement over the measures [7], thus R and A should be interpreted in a combined manner.

The aim of the present study is to investigate if ECG measurements are reliable for low sampling frequencies. In this work, the reliability of RR, RT, QT and QT_c beat-to-beat measures using low sampling frequency were quantified with respect to the measures at original rate, using the CCC and ICC, while agreement was accessed using the Information–based measure of disagreement (IBMD).

2. Data and methods

All the processing and statistical analysis was performed using the MATLAB language.

2.1. Databases

The PTB diagnostic ECG database (PTBDB) [8, 9], available at `www.physionet.org` consists in 549 records from 294 subjects. Each record contains the standard 12–lead ECG and the simultaneously recorded 3 Frank lead ECG. The signals are originally sampled at 1000 Hz, with a resolution of 0.5 μV and have variable duration (minimum 30 sec, typically around 2 min). The database includes 54 healthy controls as well as patients with different pathologies, mainly cardiac related.

A second dataset consists on 12 lead 1000 Hz Holter recordings acquired with Mortara H12+, from 26 pediatric patients (age < 18) with ABI admitted in the Pediatric Intensive Care Unit of Centro Hospitalar São João. The patients are part of a database (PICUDB) collected between 2006 and 2014, obtained under informed consent, in a project approved by the respective ethic commission and by the Portuguese data protection authority. From those 26 patients, BD has been confirmed, during the recording or at a latter moment by the usual protocol, for 11 patients, corresponding to 23 recordings; other 2 died during the first recording and 13 patients (28 recordings) survived. The first hour of each of the 53 recordings was considered in this study in a total of 53 hours. The typically recorded lead in monitoring systems, Lead II, was used.

2.2. Time series extraction

The original 1000 Hz signals in both databases were downsampled to 500 Hz and 250 Hz. A multiscale wavelet-based ECG delineator previously developed and validated [10] was used. In this method, a discrete wavelet transform is applied producing coefficients proportional to the derivative of the ECG signal smoothed at different scales. The algorithm then searches local extrema of the differentiated signals detecting and classifying the relevant slopes at different scales, according to the different spectral content of each ECG wave (e.g. for QRS complex and T wave). A multiscale threshold approach is used to locate the waveform limits.

For PTBDB each ECG lead was automatically delineated to obtain single lead based annotations (SL) for each sampling frequency considered. Multilead based annotations were obtained from leads I, III, V1-V6 as the median mark of the 8 SL annotators for the R peak, while for QRS onset [T end] the boundary is assumed as the first [last] SL based location with 2 other SL locations within 6 ms [12 ms] tolerance [10]. Lead II of PICUDB Holter signals for each sampling frequency considered was also automatically delineated using he strategy in [10]. The RR intervals for each beat were calculated considering the time intervals between the dominant QRS main wave (QRSp) in the file, RT values were taken as the time intervals from QRSp and T wave peak and QT values were taken as the time intervals from QRS complex onset and T wave end. The RR and QT time series were aligned across the sampling rates to ensure that the measures with respect to the same beat are compared and QT$_c$ was obtained from Bazzett correction formula as $QT_c = QT/\sqrt{RR}$.

2.3. Reliability estimation methods

In this work, the reliability (R) is measured considering Lin's concordance correlation coefficient (CCC) and intraclass correlation coefficients (ICC).

Let X_i $(x_1, ..., x_n)$ and Y_i $(y_1, ..., y_n)$ be the values obtained by the two methods in i^{th} subject with means μ_X and μ_Y, variances σ^2_X and σ^2_Y and covariance σ_{YX}. The Lin concordance correlation coefficient is defined by equation (1), combining measures of both precision and accuracy to determine how far the observed data deviate from the line of perfect concordance line at 45° on a square axis scatter plot.

$$\text{CCC} = \frac{2\sigma_{YX}}{(\mu_Y - \mu_X)^2 + \sigma^2_Y + \sigma^2_X} \tag{1}$$

Alternatively, R can be quantified as a measure of the amount of the total variance attributable to true differences and expressed as the ratio of true score variance to total variance. According to [6], the intraclass correlation coefficient (ICC), in the case in which each subject is evaluated by each method, is defined by equation (2), where V_{true} (true variance) and V_{total} (total variance) are considered as usually in ANOVA and summarized in Table 1.

$$\text{ICC} = \frac{V_{\text{true}}}{V_{\text{total}}} \tag{2}$$

Table 1. ICC from Two–way ANOVA: F values are the variance ratio distributions, for mean square (MS) due to patients (MSP) or to methods (MSR) or to error (MSE). V_{f} stands for the variance associated to the factor f.

Source	Variance	F-ratio	ICC
Patients/beats	V_{patients}	$\frac{\text{MSP}}{\text{MSE}} = \frac{V_{\text{patients}}}{V_{\text{error}}}$	$\text{ICC} = \frac{V_{\text{true}}}{V_{\text{total}}}$
Methods	V_{methods}	$\frac{\text{MSR}}{\text{MSE}} = \frac{V_{\text{methods}}}{V_{\text{error}}}$	
Interaction	V_{true}	$\frac{\text{MSI}}{\text{MSE}} = \frac{V_{\text{true}}}{V_{\text{error}}}$	
Error	V_{error}		
Total	V_{total}		

The indexes CCC and ICC are interpreted as absolute correlation coefficients, ranging from 0 (no agreement) to 1 (perfect agreement). Since it's not established which one is the best, we use both coefficients in this work. In addition, the ICC has the advantage of adapting to the case in which each beat (subject) is evaluated by each method [6].

A new information-based measure of disagreement (IBMD) was proposed in [7] to cope with the fact that no available strategy allowed to compare the degree of agreement, without the influence of the reliability. This approach provides a method based on Shannon's entropy to compare the degree of disagreement among different populations, as it is here the case by considering several sampling frequencies. The information based measure of disagreement (IBMD) between non-negative measurements obtained by the methods Y and X, is defined as:

$$\text{IBMD} = \frac{1}{n}\sum_{i=1}^{n}\log_2\left(\frac{|x_i - y_i|}{max\{x_i, y_i\}} + 1\right) \qquad (3)$$

with the convention $\frac{|0-0|}{max\{0,0\}} = 0$.
This coefficient equals 0 when the observers agree (no disagreement), ie when $x_i = y_i$. In this case there is no information in the differences between methods X and Y. The more the measurements from X and Y disagree, the more the amount of information in the differences between observers increases, thus IBMD increases towards to 1. The agreement (A) can be quantified as $\text{A} = 1 - \text{IBMD}$.

3. Results and discussion

Reliability (R) and agreement (A) were evaluated over PTBDB 549 recordings considering the mean intervals per file in Lead II (M), one normal beat per recording for each of the 15 leads (SL) and for the multilead based intervals (SLR); over 53 PICUDB 1–hour recordings for 50 random selected beats (G0), considering together all 50 beats*53 recordings (G1) and one arbitrary beat per file (G2).

The conventional interpretation was used: Excellent (R$>$ 0.9, A$>$ 0.9), Good (R$>$ 0.7) and Poor (otherwise). For each time series studied and sampling rate value were performed 72 estimations: 15SL+SLR+M+53N+G1+G2 (for N files in G0). Regarding CCC and ICC the results are summarized in Table 2 for all data. Since CCC and ICC quantify the same, we obtained the expected concordance and equivalent results were found for more than 97% of the comparisons, with lower R by ICC. According to ICC, excellent reliability was found at 500 HZ for 83% , 79%, 81% and 78% of the cases for RR, RT, QT and QT_c, respectively, with slightly lower results at 250 HZ: 68% , 50%, 50% and 58%. With respect to agreement, IBMD$<$ 0.1 (excellent) was found in all PTBDB and PICUDB cases, except at 2 PICUDB cases for RT, QT and QT_c. Thus, in spite that both R indexes quantifying reliability affected by agreement, any reductions should be explained by lower reliability and not by agreement changes, since the agreement keeps high. The high reliability, low disagreement and high CCC vs ICC equivalence are well illustrated by the distributions of the indexes across all beats considered from PICUDB (G1) presented in Figure 1, with shaded areas limiting the values for interpretation.

Table 2. Summary of results for CCC and ICC: E, G and P stand, respectively for Excellent (R$>$ 0.9), Good (R$>$ 0.7) and Poor (otherwise); N stands for PICUDB files in G0.

	PTBDB			
500Hz	RR	RT	QT	QT_c
E	M, 11SL,SLR	M, 15SL, SLR	M, 15SL, SLR	M, 15SL, SLR
G	4 SL	-	-	-
P	-	-	-	-
	PTBDB			
250Hz	RR	RT	QT	QT_c
E	8SL	M, 8SL, SLR	M, 13SL, SLR	SLR, 13SL
G	M, 7SL, SLR	7SL	2SL	M, 2SL
P	-	-	-	-
	PICUDB			
500Hz	RR	RT	QT	QT_c
E	45N, G1–2	40N	41N	39N
G	3N	9N, G1–2	8N, G1–2	9N, G1–2
P	5N	4N	4N	5N
	PICUDB			
250Hz	RR	RT	QT	QT_c
E	39N, G1–2	26N	19N, G1–2	27N, G2
G	8N	16N, G1–2	20N	16N, G1
P	6N	11N	14N	10N

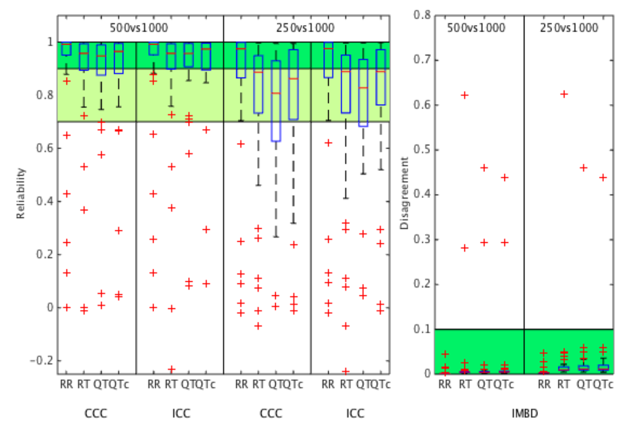

Figure 1. Boxplots with results for beats from all 53 patients (G1): Reliability and disagreement (IBMD) for RR, RT, QT and QT_c series, Excellent (R,A$>$ 0.9) in darker green, Good (R$>$ 0.7) in lighter green.

As expected, reliability is slightly lower at 250 Hz than at 500 Hz. Strange results were found for G1-2, with higher R values for QT and QT_c at 250 Hz than at 500 Hz. This can be due to the high fitting of the delineation method for T wave end detection at 250 Hz, while at 500 Hz, adapted filters are used [10]. Nevertheless, this is not the case in the majority of the comparisons by recording, thus should not represent an advantage for the clinical analysis. The very good T wave delineation can also explain the small reliability differences between RT and QT, with even better results for QT for G1–2 at 250 Hz (Table 2).

In general the low disagreement and high reliability found between the time series measures at 1000 Hz and lower sampling frequencies indicate that the measurements can still accurately taken, even when ECG of high resolution is not available. An overestimation of the QT interval when decreasing the sampling rate below 300 Hz was described by Risk et al [11]. In that work the effect of the sampling rate was modeled as an exponential decay function and the lower sampling frequency lower limit taken as the point in which the model reaches the asymptotic value. Values of 290 Hz and 303 Hz were reported respectively for QT and QT_c. As in [11] a dedicated unvalidated strategy was used to delineate and measure RR and QT intervals, we assume that delineation errors could have reduced the performance at lower sampling frequencies. Moreover, their conclusions were based in the assumption of a decaying model and no statistical comparison was performed. On the other hand, at the present study, we used a previously validated delineation system that presented good performance for QT measuring [12] and investigated if ECG measurements are still reliable for low sampling frequency using well established reliability indexes.

4. Conclusions

Our main goal was to study the reliability of the RR, RT, QT and QT_c beat-to-beat measures using low sampling frequency, tipically available at bedside monitoring systems. The statistical equivalence was evaluated using both reliability and agreement indexes, through CCC, ICC and IBMD. Results indicated that even at the lowest sampling rate 250 Hz, all time interval series are still reliable compared with the values at 1000 Hz, with small measurement errors in comparison to the true differences between subjects. Therefore low sampling frequency ECG derived intervals can be considered consistent and used as reliable.

Acknowledgements

Research partially supported by CMUP (UID/MAT/-00144/2013), funded by FCT (Portugal) with national (MEC) and European structural funds through European Regional Development Fund (FEDER), under the partnership agreement PT2020, by project TEC2013-42140-R from the Ministerio de Economia y Competitividad (MINECO, Spain) with FEDER, and by Grupo Consolidado BSICoS (T-96) from DGA (Aragón, Spain) and European Social Fund (EU). The authors acknowledge the support from Mortara Instr Inc., represented in Portugal by CardioSolutions, Lda.

References

[1] Mazzeo A, La Monaca E, Di Leo R, Vita G, Santamaria L. Heart rate variability: a diagnostic and prognostic tool in anesthesia and intensive care. Acta Anaesthesiologica Scandinavica 2011;55(7):797–811.

[2] Rocha AP, Almeida R, Silva MJ, Silva ME. Long-term hrv in critically ill pediatric patients: coma versus brain death. In Computing in Cardiology Conference (CinC), 2014. IEEE, 2014; 89–92.

[3] Almeida R, Silva MJ, Rocha AP. Exploring qt variability dependence from heart rate in coma and brain death on pediatric patients. In Computing in Cardiology Conference (CinC), 2013. Sept 2013; 61–64.

[4] Keating J, Matyas T. Unreliable inferences from reliable measurements. The Australian journal of physiotherapy 1998;44(1):5.

[5] Bartlett J, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound in Obstetrics Gynecology 2008; 31(4):466–475.

[6] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological bulletin 1979;86(2):420.

[7] Costa-Santos C, Antunes L, Souto A, Bernardes J. Assessment of disagreement: a new information-based approach. Annals of epidemiology 2010;20(7):555–561.

[8] Bousseljot R, Kreiseler D, Schnabel A. Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet. Biomedizinische TechnikBiomedical Engineering 1995; 40(s1):317–318.

[9] Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov P, Mark R, Mietus J, Moody G, Peng CK, Stanley H. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215–e220.

[10] Martínez JP, Almeida R, Olmos S, Rocha AP, Laguna P. A wavelet-based ecg delineator: evaluation on standard databases. Biomedical Engineering IEEE Transactions on 2004;51(4):570–581.

[11] Risk M, Bruno J, Soria ML, Arini P, Taborda R. Measurement of qt interval and duration of the qrs complex at different ecg sampling rates. In Computers in Cardiology, 2005. IEEE, 2005; 495–498.

[12] Martinez J, Almeida R, Olmos S, Rocha A, Laguna P. Stability of qt measurements in the ptb database depending on the selected lead. In Computing in Cardiology Conference (CinC), 2006. Sept 2006; 341–344.

Address for correspondence:

Paulo Sousa, Rute Almeida, Ana Paula Rocha
CMUP & Faculdade de Ciências da Universidade do Porto
Rua do Campo Alegre, 687, 4169–007 Porto, Portugal
E-mail: paulo.sousa, rbalmeid, aprocha@{fc.up.pt}