Detection Quality Indices for Improved Heart Beat Assessment in Non-Invasive Fetal ECG

Jonas Weiß, Hagen Malberg, Martin Schmidt

TU Dresden, Institute of Biomedical Engineering, Dresden, Germany

Abstract

In this work, we establish the novel concept of detection quality indices (DQI) for the assessment of non-invasive fetal heartbeat detections. We grouped the DQIs in four categories concerning the detection properties they address and used these DQIs as features to classify correct detections and two detection errors with a decision tree. Our classifier was able to distinguish between correct detection and two detection errors (f1 = 0.90). By excluding the detection errors, we were able to improve the error of time-domain heart rate variability parameters and approximated entropy compared to manual detection by 35 percentage points on average in a first clinical test setting.

1. Introduction

Non-invasive fetal ECG (fECG) is a promising technology for long-term monitoring of pregnancies and thus the early detection of fetal abnormalities. One challenge of most importance is the detection of fetal QRS complexes. Due to many works in this field, several high-performance QRS detection algorithms exist in the literature. But still, there are some detection errors, that occur regularly as a result of temporal changes in signal quality [1]. A common strategy is to optimize the QRS detector in the way that detection errors occur less often. In this work, we chose an alternative approach by classifying the QRS detentions as valid or as one of two types of errors. This was implemented by using designed features and a decision tree to understand the classifier's behaviour better. Later we approve the clinical benefit of this method on our ongoing Fetal Autonomic Cardiovascular rEgulation (FACE) study.

2. Materials and Methods

In this work, we used a subset of the FACE study. Fetal QRS complexes were detected with the evolutionary fetal QRS detector (classical detector) described by Andreotti et al. [2]. This detector works with a naive Bayes approach and is highly robust against noise. However, in some cases, it fails to detect the fetal QRS complexes but instead detects peaks introduced by noise or residuals of the maternal QRS complexes. To take this into account we categorized detections in: valid (*fQRS*), noise peak (*noise*), and maternal QRS complex residual (*mRES*).

For the classification, we used indices that measure the quality of a single detected beat and referred them as Detection Quality Indices (DQI). The idea derives from the work on Signal Quality Indices (SQI), which are widely used, in the field of ECG and fECG [2–4]. SQIs are usually used to assess the quality of a signal. All DQIs used in this work a summarized in Table 1. To the best knowledge of the authors, only two SQIs for fECG are currently available that can measure the quality of a single beat detection (cSQI and xSQI [2–4]).

If the detections are categorized, the invalid detections can be excluded. In a further step, an enhanced fetal QRS detector could be designed based on the classifier's metric. In this work, we focus on the filtering approach, by exclusion.

2.1. Signal Quality Indices

xSQI measures the e**x**travagance by comparing a series of fQRS peaks to the power of the surrounding signal [3,4]. We chose a window of 60 ms as relevant for fetal QRS complex to capture the full QRS complex independent of the week of gestation (WOG) [5]. By comparing every beat with the surrounding 350 ms. We referred to this parameter as xDQI.

cSQI characterizes the **c**onformity of a beat series compared to an averaged fQRS template achieved by Pearson correlation [2]. In this work, we defined cSQI considering a single beat detection as $cDQI_m$.

2.2. Novel Detection Quality Indices

aDQI defines the **a**greement of maternal heart rate and fetal heart rate as the ratio at the time of every detection. Hence, the heart rates were calculated from the RR intervals for the full measurement and were linear interpolated.

$cDQI_{TDE}$ is based on $cDQI_m$. The template of $cDQI_m$ is naive. Therefore, we exchanged the way of building the

template by the procedure described in [6]. This procedure is based on time delay estimation to shift the QRS detections to temporally adjust annotations. The used Improved Woody's Method displaces annotations in a way that the correlation of annotations surrounding signal segments is maximized. This approach considers not only the peak of the QRS complexes but also the morphology[1] and thus the template is more robust to small detection errors. We also used the first and second derivative of the template and its correlation with the current QRS complex resulting in two additional DQIs. For reasons of comparison , we did this procedure not only with the $cDQI_{TDE}$, but also with the $cDQI_m$. The derivatives are labelled as $cDQI_{TDE}^0$ ($cDQI_m^0$) for the initial template, $cDQI_{TDE}^1$ ($cDQI_m^1$) for the first derivative and so on.

dDQI characterizes the change in fetal heart rate, calculated by the second **d**erivative of the heart rate time series for each fetal QRS complex. The idea is, that false detections should be followed by an atypical change in the fetal heart rate.

nDQI characterizes the **n**umber of fetal QRS detections overlapping with maternal QRS complex detections. A similar concept was used in the iSQI [3] and miSQI [2]. iSQI is defined as the percentage of beats detected in one lead, also detected in another lead. miSQI is defined as $1-$ iSQI, comparing fQRS complexes and maternal QRS-complexes (mQRS). This definition is depending on fetal and maternal heart rate. Thus, we defined nDQI as the cohens-κ of the agreement of mQRS and fQRS in a certain window [7]. The agreement was defined as an fQRS complex in an acceptance-interval to an mQRS complex. nDQI is defined in Equation 1.

$$nDQI = \frac{p_{m,f} - p_0}{1 - p_0} \qquad (1)$$

$p_{m,f}$ represents the fraction of agreement. The probability of random agreement p_0 is defined by the sum of all acceptance intervals divided by the total window length. In this work, we used an acceptance-interval of 25 ms and a window of 10 seconds.

2.3.   Data

The used data originate from the FACE study which is currently in progress at the University of Leipzig Medical Center in cooperation with the HU Berlin and the TU Dresden. The study was approved by the committee of ethics of the University of Leipzig Medical Center (357/17-ek). The fetal ECG signal was recorded as described by Andreotti et. al [4]. Our measurement protocol included paced slow

---

respiration, paced fast respiration, and an introduced loss in blood pressure. The measurement begins with a 10 min interval of rest. After all interventions, there is a 5 min interval of rest.

As test and training dataset for the classifier, we used 14 measurements of 11 healthy pregnant women. At the week of gestation 22 to 28 (mean 24.6) and age 20 to 40 (mean 30.5) that revealed detection errors. We carefully assigned the previously defined detection categories to automated QRS detections. We randomly assigned the measurements to training and test set and picked random beats for training (n = 8,796) and test set (n = 3,768), with equal size for each category. The decision tree was trained in *scikit-learn* with a maximum depth of five and a minimal leave size of 30 to prevent overfitting [8]. The method was further evaluated by comparing the classical and the filtering approach (augmented detector) with manual annotations. We used the last 5 minutes of the rest interval of two measurements of one healthy subject (WOG 22 and 25, age 24) and evaluated the heart rate variability (HRV) [9].

2.4.   Statisics

To evaluate the DIQs, we tested the difference of indices by a two-sided t-test and calculated the effect size as the Spearman correlation coefficient (ρ), after ensuring for normal distribution. The effect sizes can be used, as a measure for the information expressed by this index. An absolute ρ of 0 would indicate non-effect, an absolute effect of 1 indicates a strong effect. The feature importance was assessed by the Gini coefficient [10].

Table 1. Overview of all used DQIs in this work.

DQI	description	origin
aDQI	fHR/mHR for every fetal Beat	this work
$cDQI_{TDE}^0$	conformity to QRS template	this work
$cDQI_{TDE}^1$	conformity to first derivative of QRS template	this work
$cDQI_{TDE}^2$	conformity to second derivative of QRS template	this work
$cDQI_m^0$	conformity to QRS template	[2]
$cDQI_m^1$	conformity to first derivative of QRS template	this work
$cDQI_m^2$	conformity to second derivative of QRS template	this work
dDQI	second derivative of fHR	this work
nDQI	fQRS complexes close to mQRS complexes	this work
xDQI	extravagance of QRS complex	[3,4]

3.   Results

3.1.   Evaluation of the DQIs

The results of the correlation analysis between the DQIs are shown in Figure 1. dDQI and nDQI contained no strong correlation with other DQIs. Thus they contain more unique information about the detection or none at all. The cDQIs were highly correlated.

---

[1]TDE could also be used to refine the initial fQRS detections, but in this work we do not use the displacement information.
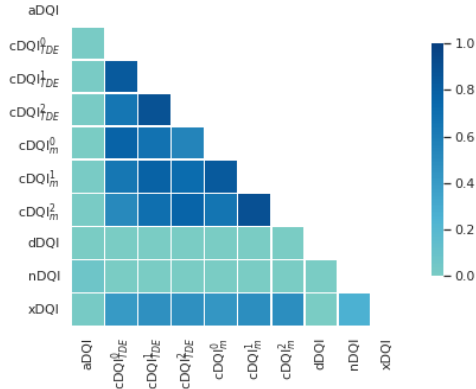
Figure 1. Spearman correlation matrix of the DQIs for all values from the test and training dataset.

To characterize the value of every DQI for each category we considered the spearman correlations between all categories. The results are shown in Table 2. All DQIs except dDQI and xDQI showed a p-value smaller than 0.001 and thus show a significant difference between the groups.

aDQI yielded exceptional high effect sizes (up to $\rho = -0.83$). This suggested that DQI is a measure of fQRS detections similar to mQRS detections. It was therefore plausible that this DQI is capable of distinguishing between *fQRS* and *mRES* ($\rho = 0.47$). We conducted a deeper examination of the data and found that the used detector indeed tends to favour *mRES* if no *fQRS* is found.

xDQI yield their highest effect by comparing fQRS and mQRS, even if the effect sizes are still low (-0.17).

The two classes of cDQIs showed similar power. $cDQI_m$ yielded higher effect sizes for *fQRS vs noise* and *noise vs mRES* and *fQRS vs mRES*, but $cDQI_{TDE}$ yield only 0.05 smaller effect sizes. Hence, we considered both categories of DQIs as beneficial. The results are in accordance, with the categories we proposed earlier (Figure 1).

Hence, we grouped the DQIs in four categories: information about the ration of fetal-maternal heart rate (aDQI), the similarity of fQRS and mQRS (nDQI), morphological information (cDQIs) and extravagance of a detection (xDQI). dDQI does not contain any relevant information.

3.2. Evaluation of the Classifier

Our results showed that the developed classifier can distinguish between the given category (accuracy, precision, recall and f1 > 0.90 for all categories, cohens $\kappa = 0.85$). We also checked the feature importance of the DQIs. nDQI has the highest importance (0.76). This was plausible, due to the previously discussed tendency of the detector to favour *mRES* if no *fQRS* is found. $cDQI_m$ was also important (0.17). The novel $cDQI^1_{TDE}$ and the xDQI had also

Table 2. Effect size as the Spearman-correlation coeffcient (ρ) and the results of testing for non correlation indicated by a two tailed p-value. *** stands for $p < 0.001$.

DQI	fQRS vs noise		fQRS vs mRES		noise vs mRES	
	p-value	ρ	p-value	ρ	p-value	ρ
aDQI	***	-0.83	0.02	0.03	***	0.47
$cDQI^0_{TDE}$	***	-0.24	***	-0.43	***	-0.11
$cDQI^1_{TDE}$	***	-0.49	***	-0.59	***	-0.08
$cDQI^2_{TDE}$	***	-0.56	***	-0.51	***	-0.06
$cDQI^0_m$	***	-0.28	***	-0.48	***	-0.12
$cDQI^1_m$	***	-0.51	***	-0.61	***	-0.09
$cDQI^2_m$	***	-0.60	***	-0.55	***	-0.07
dDQI	0.91	0.00	***	-0.07	0.09	-0.02
nDQI	***	0.40	***	0.12	***	-0.09
xSQI	0.47	-0.01	***	-0.17	***	-0.06

significant importance, but their absolute impact is much smaller (0.02 and 0.04). All other DQIs yielded a Gini coefficient less than or equal 0.01.

3.3. Comparison with Manual Annotation

We compared two measurements of a healthy pregnant subject that we annotated manually. The augmented detector found 97 % or more of the manually annotated beats and excluded up to 2 % of false detections. In both measurements, the excluded beats where predominant classified as *noise*. Up to 3 % of the beats annotated manually were not detected by the augmented detector.

A comparison of HRV parameters calculated from the resting periods of each measurement are shown in Table 3 for all three fQRS detection methods. The minor exclusions of beats had a clear impact on the error of HRV parameters. To compare the HRV parameters we calculated the relative difference of each HRV parameter for the classical and the augmented detections compared to the manual annotations and found a decrease for the following HRV parameters: SDSD (0.50 to 0.15), SDNN (0.80 to 0.15), and Approximated Entropy (ApEn) (0.50 to 0.10). The triangular index (TRI) was robust to the exclusion of beats and yields consistent results for all detection methods. This is a promising result since we just excluded up to 2 % of invalid beat detections and not take further steps to find QRS complexes in the areas of exclusion.

Table 3. Comparison of HRV parameters calculated form RR intervals of the classical detector, the augmented detector, and the manual anotations.

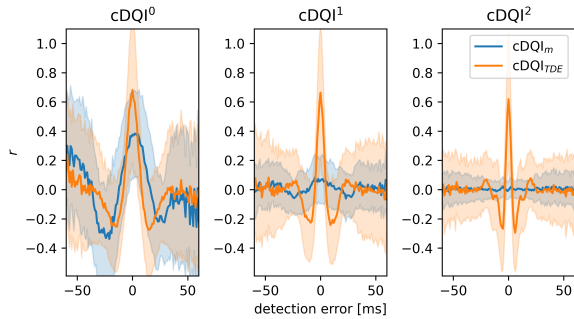|	classical		augmented		manual	
WOG	22	25	22	25	22	25
SDSD	0.02	0.03	0.07	0.38	0.06	0.33
SDNN	0.02	0.02	0.05	0.38	0.06	0.24
TRI	1.83	1.70	1.84	1.67	1.84	1.69
ApEn	0.44	0.41	0.38	0.25	0.40	0.16

Figure 2. Comparison of mean and standard deviation of $cDQI_m$ and $cDQI_{TDE}$ for introduced detection errors (μ = 0 ms, σ = 50 ms, n = 37,000).

4.   Discussion

In this work detection quality indices (DQIs) were introduced and established. We categorized the DQIs in four independent classes of information: ration of fetal-maternal heart rate, the similarity of fQRS and mQRS detections, the extravagance of detection, and morphological information. By training a decision tree, we were able to distinguish between correct detected fQRS complexes and two detection errors. We found that the classical detector could be improved by a better evaluation between fQRS and mQRS.

Correlation analysis indicated that all cDQIs are highly correlated. Thus we compared all by a synthetic dataset, generated with fecgsyn [11]. The comparison results in a random detection error, shown in Figure 2 because we expect the TDE to have a significant impact on random detection errors. In $cDQI^1$ and $cDQI^2$, the position of the true QRS complex is more clear for the $cDQI_{TDE}$. We conclude that the novel $cDQI_{TDE}$ estimates the exact position of the fQRS complex better than the $cDQI_m$.

By comparing manual QRS annotation and augmented detection for two fECG measurements of pregnant woman, we found the first evidence of benefits for the calculation of HRV parameters. Therefore, we conclude that DQIs are useful to understand the behaviour of a QRS detector better and thus can improve the quality of QRS detections.

Further studies should take into account other detectors, and examine the effect on HRV parameters further. Furthermore, the indices introduced in this work could be used in the design of an enhanced fetal QRS detector.

Acknowledgments

References

[1] Weiss J, Malberg H, Schmidt M. Non-invasive foetal ECG signal quality assessment by experts and non-experts. In 2020 11th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO). IEEE, July 2020; .

[2] Andreotti F, Gräßer F, Malberg H, Zaunseder S. Non-Invasive Fetal ECG Signal Quality Assessment for Multichannel Heart Rate Estimation. IEEE Transactions on Biomedical Engineering December 2017;64(12):2793–2802.

[3] Li Q, Mark RG, Clifford GD. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. Physiological Measurement December 2007;29(1):15–32.

[4] Andreotti F, Riedl M, Himmelsbach T, Wedekind D, Wessel N, Stepan H, Schmieder C, Jank A, Malberg H, Zaunseder S. Robust Fetal ECG Extraction and Detection from Abdominal Leads. Physiological Measurement August 2014; 35(8):1551–1567.

[5] Brambati B, Pardi G. The intraventricular conduction time of fetal heart in uncomplicated pregnancies. BJOG An International Journal of Obstetrics and Gynaecology November 1980;87(11):941–948.

[6] Schmidt M, Baumert M, Porta A, Malberg H, Zaunseder S. Two-dimensional warping for one-dimensional signals—conceptual framework and application to ECG processing. IEEE Transactions on Signal Processing November 2014;62(21):5577–5588.

[7] Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement April 1960; 20(1):37–46.

[8] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 2011; 12:2825–2830.

[9] Vest AN, Poian GD, Li Q, Liu C, Nemati S, Shah AJ, Clifford GD. An open source benchmarked toolbox for cardiovascular waveform and interval analysis. Physiological Measurement October 2018;39(10):105004.

[10] Gini C. Variabilitàe Mutuabilità. Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche. Bologna: C. Cuppini, 1912.

[11] Behar J, Andreotti F, Zaunseder S, Li Q, Oster J, Clifford GD. An ECG simulator for generating maternal-foetal activity mixtures on abdominal ECG recordings. Physiological Measurement July 2014;35(8):1537–1550.

Address for correspondence:

Jonas Weiß
TU Dresden, Institute of Biomedical Engineering
01062 Dresden, Germany
jonas.weiss@tu-dresden.de