KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Moeyersons J., Testelmans D., Buyse B., Willems R., Van Huffel S., Varon C. (2018), **Evaluation of a Continuous ECG Quality Indicator Based on the Autocorrelation Function** In Proceedings of the 45th Annual Computing in Cardiology Conference (CinC 2018), Maastricht, The Netherlands, Sep. 2018.
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	/
Journal homepage	https://www.cinc2018.org/
Author contact	Jonathan.moeyersons@kuleuven.be 0479427028
Abstract	
IR	Na.

Evaluation of a Continuous ECG Quality Indicator Based on the Autocorrelation Function

Jonathan Moeyersons[1,2], Dries Testelmans[3], Bertien Buyse[3], Rik Willems[4], Sabine Van Huffel[1,2], Carolina Varon[1,2]

[1]KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Leuven, Belgium
[2]Imec, Leuven, Belgium
[3]Department of Pneumology, UZ Leuven, Leuven, Belgium
[3]Department of Cardiovascular Sciences, UZ Leuven, Leuven, Belgium

Abstract

Most electrocardiogram (ECG) signal quality assessment algorithms focus on a two- or multi-level classification. However, it could be argued that signal quality would more naturally occupy a continuum of quality values. Therefore, in previous work we created a continuous quality assessment algorithm based on the autocorrelation function (ACF). This paper evaluates this algorithm on a simulated dataset with five noise levels and known signal-to-noise ratios (SNR).

The simulated data was created by selecting clean ECG segments of a polysomnographic dataset with an in-house quality algorithm, and adding calibrated amounts of two types of realistic ECG noise from the MIT-BIH Noise Stress Test Database (NSTDB). Both Electrode Motion (EM) and Movement Artefacts (MA) were considered.

Using only three features and a binary training set, we have shown significant quality decreases per noise level for both types of added noise. Despite this finding, also significant intra-level differences were observed, indicating a change in response according to the type of noise. Adding other features might help to converge the quality scores.

By presenting the users with a continuous quality score, they are given the possibility to define the preferred level of quality according to the study objective.

1. Introduction

Electrocardiogram (ECG) recordings are often contaminated with large amounts of noise and artefacts [1]. The presence of these contaminating factors could reduce the diagnostic capabilities of the ECG and degrade the performance of the most sophisticated signal processing algorithms [2]. Since these algorithms are usually designed to operate on signals whose quality has already been verified by an expert, an indication of the signal quality is necessary [3].

Most of the earlier work on quality indication is restricted to binary classification: clean or contaminated, acceptable or unacceptable. However, the acceptance stringency of this binary classification might vary significantly according to the study objective [3]. For instance, heart rate variability studies require a low acceptance stringency, since only QRS-complexes need to be accurately detected, while studies that investigate waveform morphology changes require a high acceptance stringency.

As an improvement on the binary classification, some authors proposed to use multiple discrete noise levels. Vaglia et al. defined three signal quality bins: low, average and good quality. Signals in the low-quality bin will need manual reading, signals with average quality will need cardiologist over-read and good quality signals will not need any further review process [4]. In more recent work, Li et al. presented a five level signal quality classification algorithm: clean, minor noise, moderate noise, severe noise and extreme noise [5].

Redmond et al. stated that while discrete quality labels facilitate the creation of an expert labeled gold standard training set, it could be argued that signal quality would more naturally occupy a continuum of quality values [3]. Therefore, starting from a previously proposed binary quality indication algorithm, based on the autocorrelation function (ACF) and a RUSBoost classifier, we suggested to use the weighed sum of the posterior probability of the clean class as a continuous indication of the signal quality [6].

In this paper, the mentioned continuous signal quality indicator is evaluated on a simulated dataset with five noise levels and known signal-to-noise ratios (SNR).

2. Materials & Methods

2.1. Clean data

The quality assessment algorithm was evaluated on a large subset of the ECG signals of a polysomnographic dataset [7]. Briefly, the dataset consists of 16 single-lead (lead II) ECG recordings, originating from 16 different patients. A total amount of 152 hours and 12 minutes of signal was acquired with a sampling frequency of 200 Hz.

Since no ECG signal quality annotations of this dataset exists, an in-house quality estimation algorithm was applied to locate the cleanest epochs [7]. The recordings were firstly segmented in 10 seconds epochs and hereafter the cleanest epochs were selected. Stringent thresholds were used to ensure the validity of this approach. This resulted in 49313 clean 10 seconds epochs.

2.2 Additive noise

The PhysioNet noise stress test database (NSTDB) was used to generate noisy records [8]. The database contains samples of three types of noise: electrode motion (EM), baseline wander (BW) and muscle artefact (MA). Only EM and MA were considered, since baseline wander is usually not a cause for erroneous R-peak detection.

Like the clean recordings, the noisy recordings were segmented in epochs of 10 seconds. For each clean ECG epoch a noisy epoch was randomly selected and a calibrated amount of this noise epoch was added to the clean epoch.

2.3. Signal noise levels

The same signal noise levels as described in [5] were used.

- Level 0 (clean): An outstanding recording with no visible noise or artefact
- Level 1 (minor noise): A good recording with transient artefact or low-level noise that does not interfere with interpretation or recognition of P, T or atrial flutter waves
- Level 2 (moderate noise): An adequate recording that can be interpreted with confidence despite visible and obvious flaws, but does not interfere with the identification of QRS complexes or ventricular flutter waves
- Level 3 (severe noise): A poor recording that may be interpretable with difficulty. Noise interferes with QRS or ventricular flutter recognition
- Level 4 (extreme noise): An unacceptably poor recording that cannot be interpreted with confidence because of significant technical flaws.

The SNR of the resulting signals was defined as described in [9] and the different SNR levels are shown in Table 1.

Table 1: The required SNR for each type of added noise to create the same noise level. Adapted from Li et al. [5].

Noise level	Description of noise	SNR levels (dB)	
		EM	MA
Level 1	Minor	6	12
Level 2	Moderate	0	6
Level 3	Severe	-6	0
Level 4	Extreme	-12	-6

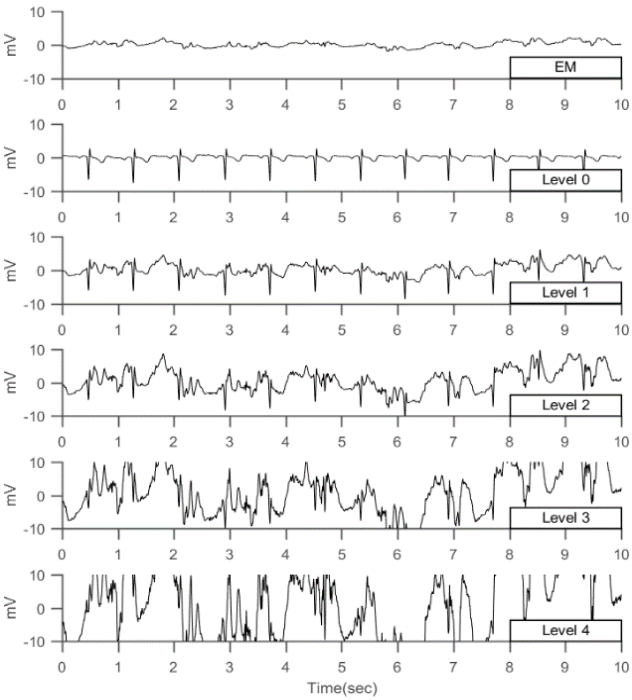

Figure 1: The diagnostic quality of the ECG signal decreases when the level of electrode motion (EM) noise increases.

This procedure resulted in a total of 98626 segments for each noise level. An example of a random EM noise sample and the different noise levels is displayed in Figure 1.

2.3. ECG quality assessment

The ECG signal quality assessment tool based on the weighed sum of the posterior probability of a RUSBoost classifier was previously described in [6]. Briefly, it consists of three steps: pre-processing, feature extraction and quality indication.

In the pre-processing step the easy-to-remove artefacts, such as baseline wander and high-frequency noise, are removed by means of a zero-phase, band pass Butterworth filter with cut-off frequencies at 1 Hz and 40 Hz. This filter effectively removes the aforementioned artefacts and

preserves the morphological structure of the ECG.

The feature extraction step consists of three smaller steps: signal segmentation, signal characterization and feature extraction. First, the 10 seconds signal is segmented in epochs of five seconds with an overlap of 80%. This results in six epochs per signal. Hereafter, each epoch is characterized by its autocorrelation function (ACF). This characterization facilitates the separation of clean and contaminated segments, since the ACF of a clean quasiperiodic signal is significantly different from a contaminated one [7]. Finally, three features are derived from the entire set of ACFs: the first (local) minimum, the maximum amplitude at 35ms and a similarity feature.

These features, all calculated from the ACF, are fed to a RUSBoost classifier [10] which is trained on a binary class training set. The continuous quality score is generated by performing a weighed sum of the posterior probability of the clean class.

A more elaborate explanation can be found in [6].

3. Results and Discussion

A clean dataset was created by applying an in-house quality estimation algorithm on the ECG signals of a polysomnographic dataset. An overall quality level of 100% (25th percentile=100% and 75th percentile=100%) was obtained for these clean signals (Figure 2). This indicates that both quality indication algorithms define clean signals the same way and that, most probably, the selected signals were indeed noise free.

Two types of realistic ECG noise, EM and MA, at different SNR levels from the NSTDB were added to create a simulated dataset. The quality values of the EM and MA noise significantly decreased with the increasing noise levels ($p<0.01$). Due to the non-normality, which was tested with a Lilliefors test, the quality values were compared by a Kruskal-Wallis test.

Additionally, the inter-noise quality values also differed significantly ($p<0.01$). This indicates that the quality indication algorithm responds differently according to the type of noise. Although the SNR is defined in a way that the noise level is the same, the shape of the noise types will remain different. The algorithm uses only three features which are derived from the ACF and is thus heavily influenced by the shape of the noise. Adding other features might help to converge the quality scores.

When contaminated with EM noise, the quality values do not converge to 0% for the most heavily contaminated segments. This might be due to the similar morphology of the EM noise and the QRS complex. A comparable issue was reported by Li et al. when testing their signal quality indices [11].

The quality values of the signals contaminated with MA noise, tend to remain close to 100% until noise level 3 is reached. This could have two possible explanations.

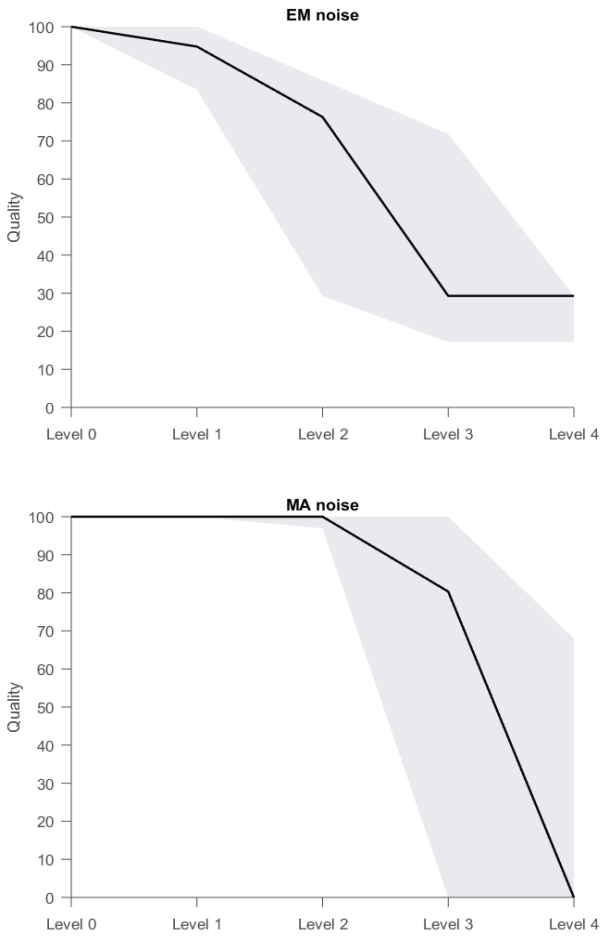

Figure 2: The quality of the EM and MA noise both significantly decrease, to a different extent, with the increasing noise level. The boundaries of the gray area indicate the 25th and 75th percentiles, and the solid line the median.

Firstly, the signal quality indication algorithm was intended for HRV analysis. Therefore the labelling of the training dataset was solely focused on accurate R-peak detection. Moody et al. have shown that MA can be tolerated at much higher levels than EM artefacts, hence higher quality scores could be expected for MA corrupted signals.

Secondly, it might be the result of the pre-processing of the signal. Before the features are extracted, the signal is band pass filtered between 1 Hz and 40 Hz. Thereby the low- and high-frequency noise is removed. Since MA noise generally consists of high-frequency signals it is possible that most of the added noise is filtered. This consideration was tested by performing only a high-pass filter instead of a band pass filter. The overall signal quality at level 2 and 3 significantly decreased.

The signal quality indicator evaluated in this study was trained on a dataset which mostly consisted of normal sinus

rhythm. Furthermore, this is also the dominant rhythm of the dataset used in this study. Previous studies have shown that signal quality classification algorithms experience a reduction in performance on an arrhythmia database when it was not explicitly retrained using signals containing arrhythmia episodes [5], [9]. Therefore, in order to further validate the algorithm, it should be tested on an arrhythmia database.

4. Conclusion

Li et al. stated that a continuous quality annotation is hard to validate since it is difficult to create a test dataset [5]. In this paper we tried to circumvent this issue by evaluating a previously developed ECG signal quality indication tool on a simulated dataset with five noise levels and known SNR's. Despite the simplicity of the algorithm, only three features were used, we have shown significant quality decreases per noise level for both types of added noise.

By presenting the users with a continuous quality score, they are given the possibility to define the preferred level of quality according to the study objective.

Acknowledgements

SV: Bijzonder Onderzoeksfonds KU Leuven (BOF): SPARKLE #: IDO-10-0358, The effect of perinatal stress on the later outcome in preterm babies #: C24/15/036, TARGID #: C32-16-00364; Agentschap Innoveren & Ondernemen (VLAIO): Project #: STW 150466 OSA + O&O HBC 2016 0184 eWatch; iMinds Medical Information Technologies: SBO-2016, ICON: HBC.2016.0167 SeizeIT; European Research Council: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (n339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information. Carolina Varon is a postdoctoral fellow of the Research Foundation-Flanders (FWO).

References

[1] G. D. Clifford, F. Azuaje, and P. E. McSharry, *ECG Statistics, Noise, Artifacts, and Missing Data*. Boston: Artech House, 2006.

[2] R. M. Rangayyan, *Biomedical signal analysis*. Johm Wiley & Sons, Inc., 2015.

[3] S. J. Redmond, Y. Xie, D. Chang, J. Basilakis, and N. H. Lovell, "Electrocardiogram signal quality measures for unsupervised telehealth environments," *Physiol. Meas.*, vol. 33, no. 9, pp. 1517–1533, 2012.

[4] M. Vaglio, L. Isola, G. Gates, and F. Badilini, "Use of ECG quality metrics in clinical trials," *Comput. Cardiol. (2010).*, pp. 505–508, 2010.

[5] Q. Li, C. Rajagopalan, and G. D. Clifford, "A machine learning approach to multi-level ECG signal quality classification," *Comput. Methods Programs Biomed.*, vol. 117, no. 3, pp. 435–447, 2014.

[6] J. Moeyersons, E. Smets, J. Morales, A. Villa, W. De Raedt, D. Testelmans, B. Buyse, C. Van Hoof, R. Willems, S. Van Huffel, and C. Varon, "Artefact detection and quality assessment of ambulatory ECG signals," *Under Rev.*, pp. 1–9, 2018.

[7] C. Varon, D. Testelmans, B. Buyse, J. A. K. Suykens, and S. Van Huffel, "Robust artefact detection in long-term ECG recordings based on autocorrelation function similarity and percentile analysis," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2012, pp. 3151–3154.

[8] G. B. Moody, W. K. Muldrow, and R. G. Mark, "A Noise Stress Test for Arrhythmia Detectors," *Computers in Cardiology*, vol. 11. pp. 381–384, 1984.

[9] G. D. Clifford, J. Behar, Q. Li, and I. Rezek, "Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms," *Physiol. Meas.*, vol. 33, no. 9, pp. 1419–1433, 2012.

[10] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost : A Hybrid Approach to Alleviating Class Imbalance," vol. 40, no. 1, pp. 185–197, 2010.

[11] Q. Li, R. G. Mark, and G. D. Clifford, "Robust heart rate estimation from multiple signal quality indices and a Kalman filter," *Physiol. Meas.*, vol. 29, no. 1, pp. 15–32, 2008.

Address for correspondence.

Jonathan Moeyersons
Kasteelpark Arenberg 10, 3001 Leuven, Belgium
Jonathan.Moeyersons@kuleuven.be