Automatic Quality Classification of Entire Electrocardiographic Recordings Obtained with a Novel Patch Type Recorder

Dorthe B. Saadi, Karsten Hoppe, Kenneth Egstrup, Poul Jennum, Helle K. Iversen, Jørgen L. Jeppesen, and Helge B.D. Sorensen

***Abstract*—Recently, new patch type electrocardiogram (ECG) recorders have reached the market. These new devices possess a number of advantages compared to the traditional Holter recorders. This forms the basis of questions related to benefits and drawbacks of different ambulatory ECG recording techniques. One of the important questions is the ability to obtain high clinical quality of the recordings during the entire monitoring period. It is thus desirable to be able to obtain an automatic estimate of the global quality of entire ECG recordings. The purpose of this pilot study is therefore to design an algorithm for automatic classification of entire ECG recordings into the groups "noisy" and "clean" recordings. This novel algorithm is based on three features and a simple Bayes classifier. The algorithm was tested on 40 ECG recordings in a five-fold cross validation scheme and it obtained an average accuracy of 90% on the test data.**

I. Introduction

Realistic long-term electrocardiogram (ECG) recordings will always contain certain amounts of artifacts including muscle artifacts, baseline wandering, electrode motions, and power line interference. This has been an unavoidable premise since the development of the first Holter recorders in the 1940s. The artifacts arise partly from normal daily life activities that neither can nor shall be avoided during the long-term recordings. The levels of noise in long-term Holter recordings have been silently tolerated, and automatic assessment of the general signal quality of Holter recordings has only obtained limited research efforts. This acceptance of the quality level is partly related to the previous lack of competitive devices. Recently, new cable-less patch type ECG recorders have reached the market [1], [2], [3]. These new devices possess a number of advantages compared to the traditional Holter recorders. These advantages include much higher patient comfort and compliance with wearing the system for extended periods of time. The extended monitoring period (up to 14 days for one device) has shown to result in the detection of more significant arrhythmias, an overall higher diagnostic yield, and a higher degree of definitive diagnosis based on the ambulatory recordings [1], [2]. However, questions concerning the long-term stability of the obtained signal quality using closely spaced recording electrodes have also been raised in the literature [4]. These new technologies thus form the basis of questions related to benefits and drawbacks of different ambulatory ECG recording techniques. This highly increases the relevance of research into areas related to automatic quantification of the clinical recording quality obtained using different techniques. Furthermore, the new technologies allow for recording of previously unknown amounts of data that need analysis. If the quality of the increased amount of data is not controlled, it might overwhelm the healthcare facilities and decrease the efficiency. This issue was also addressed by the Physionet Challenge from 2011, where participants should classify 10 seconds 12-lead ECG signals into the two groups acceptable and unacceptable [5]. In Denmark, long-term ECG recordings are analyzed by highly experienced nurses. They create the analysis reports for the referring medical doctor. These highly trained ECG technicians are accustomed to recognize disturbances as noise, and conduct the interpretation on clean data segments. However, when the general signal quality of a recording is decreased enough to interfere with the clinical interpretation and thus induces uncertainty about the analysis, the nurses write remarks of this in the analysis reports. It is highly relevant to design automatic algorithms that can mimic these subjective comments on noise levels in entire ECG recordings. The purpose of this pilot study is thus to design an algorithm that can distinguish between entire ECG recordings that are essentially noisy or essentially clean. To the best knowledge of the authors, this is the first study investigating automatic noise classification of entire ECG recordings.

This research was financially supported by The Danish Ministry of Higher Education and Science, and the Danish Market Development Fund.

D. B. Saadi is with the Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads, Bldg. 349, 2800 Kgs. Lyngby, Denmark and DELTA Danish Electronics, Light & Acoustics, Venlighedsvej 4, 2970 Hørsholm, Denmark (Phone: +45 72 19 40 00; fax: +45 72 19 40 01; e-mail: dbs@delta.dk / dorthe_bodholt@hotmail.com).

K. Hoppe is with DELTA Danish Electronics, Light & Acoustics, Venlighedsvej 4, 2970 Hørsholm, Denmark (e-mail: kh@delta.dk).

K. Egstrup is with the Department of Medical Research, OUH Svendborg Hospital, Valdemarsgade 53, 5700 Svendborg, Denmark (e-mail:Kenneth.Egstrup@rsyd.dk).

P. Jennum, H. K. Iversen and J. L. Jeppesen are with the Danish Center for Sleep Medicine, Department of Neurology, Department of Medicine, University of Copenhagen, Glostrup Hospital, 2600 Glostrup, Denmark (poul.joergen.jennum@regionh.dk, helle.klingenberg.iversen@regionh.dk, joergen.lykke.jeppesen@regionh.dk).

H.B.D. Sorensen is with the Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads, Bldg. 349, 2800 Kgs. Lyngby, Denmark (e-mail: hbs@elektro.dtu.dk).

II. Methodology

A. Data Description

We decided to use ECG data from an existing database recorded with the CE marked DELTA ePatch system that records two ECG channels with a sampling frequency of 512 Hz and an analog-to-digital converter (ADC) resolution of 12 bits [3]. The ePatch system is illustrated in Fig. 1. This pilot study includes 20 noisy recordings and 20 clean recordings. The data was extracted from a database of patients that underwent an ambulatory polysomnography at Glostrup Hospital as a part of diagnosing potential obstructive sleep

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

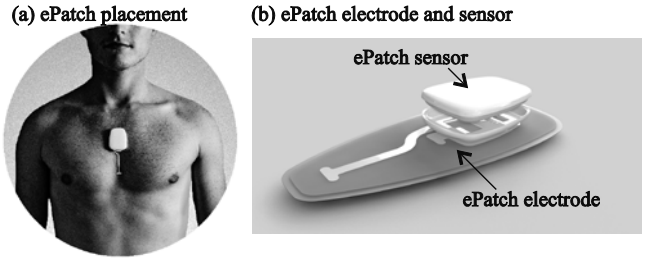

Figure 1. (a) Illustration of the ePatch system correctly placed at the sternum. (b) Illustration of the ePatch sensor and the ePatch electrode before assembly. The two ECG channels are recorded as bipolar derivations from the three skin contact points within the ePatch electrode. Modified from [3].

apnea. The demographic from each group is provided in Table I. Each of the recordings was analyzed by experienced ECG technicians using the automatic myDarwin software [6]. The noisy recordings were extracted as random recordings from the database where the ECG technician conducting the analysis included any comments on bad recording quality in either one or both channels. To ensure inclusion of both very clean and normal recordings, the 20 clean recordings were extracted as five random recordings with remarks on good quality and 15 random recordings without remarks on quality. This database thus allows for a top-level classification of entire recordings that only contains small non-disturbing amounts of noise and recordings that contain noise to an extent where the ECG technician felt enough insecure due to noise to make remarks in the analysis report. In some recordings, the system was obviously detached from the patient before the end of the data file. This is especially pronounced in the clean recordings, whereas it is difficult to judge in many of the noisy recordings. The length of all clean recordings was therefore defined by visual inspection of the recorded data.

B. Algorithm Overview

The algorithm consists of two steps: Feature extraction and classification. The algorithm output is a classification of entire two-channel ECG recordings into one of the two groups "noisy" and "clean" recordings. The algorithm applies three different features that are designed to describe some of the characteristic differences between a clean and a noisy ECG signal. The features are generally based on measuring the amount of time where the recording is noisy based on simultaneous information from both ECG channels.

C. Feature Extraction

The first feature, F_1, describes the amount of saturation in the signal. Saturation is not intended in a clean ECG signal. The feature is defined by (1), where j indicates the channel number, N is the total length of the recording in samples, x_j is the j'th ECG channel, "logical" is the value 1, when the expression is true, and 0 otherwise, "|" is the or operator, α is the maximum possible ADC value (12 bit resolution, α = 4095), and β is the minimum possible ADC value (β = 0).

$$F_1 = \frac{1}{N}\sum_{j=1}^{2}\sum_{n=1}^{N} logical\{x_j(n) = \alpha \mid x_j(n) = \beta\} \quad (1)$$

The second feature, F_2, is a measure of the general mean value of the absolute value of the raw ECG signal in non-overlapping one minute windows. In a clean ECG signal, most samples are expected to obtain a low value corresponding to the isoelectric line between the T-waves and the P-waves. A noisy segment, on the other hand, will typically contain a certain amount of samples that are significantly different from (numerical higher than) the expected isoelectric line. The mean value of a clean ECG signal is thus expected to be lower than the mean value of a noisy ECG signal with the same heart rate (HR). It is extremely important to scale the signal to attenuate the influence of the general amplitude in the recording. The amplitude can vary significantly between recordings, and even within the same recording. A scaling parameter is therefore calculated for each of the non-overlapping one minute windows. The scaling parameter was found by dividing each one minute window into 30 new equally sized non-overlapping windows. The scaling parameter was set to the median value of the maximum value in each of the 30 small windows. This scaling parameter is expected to estimate the general amplitude of the QRS complexes in the current one minute window, and is thus expected to scale the absolute value of the ECG signal between 0 and 1. The scaling is illustrated in Fig. 2(b)-(c). It is observed how this novel scaling technique allows a measurement of the noise level relative to the individual QRS amplitude for each window. For each channel, the temporary feature, $F_{2,temp,j}$ is thus calculated by (2), where m is the one minute window number, s_j is the scaling parameter, Q is the number of samples in each one minute window, and "| |" is the absolute value operator. This corresponds to the mean value of the signal in Fig. 2(c). Fig. 2(d) illustrates $F_{2,temp,1}$ for the entire duration of a noisy and a clean recording.

$$F_{2,temp,j}(m) = \frac{1}{s_j(m)\cdot Q}\sum_{q=1}^{Q}|x_j(q)| \quad (2)$$

The final feature, F_2, is then calculated as the sum of the percentage of one minute windows from each channel, where $F_{2,temp,j}$ exceeds a predefined threshold, T_2. This is defined in (3), where M is the total number of one minute windows. The threshold value was set to 0.2 by visual inspection of illustrations similar to Fig. 2(d).

$$F_2 = \frac{1}{M}\sum_{j=1}^{2}\sum_{m=1}^{M} logical\{F_{2,temp,j}(m) > T_2\} \quad (3)$$

The third feature, F_3, is a measure of the number of significant signal peaks in each of the one minute windows. The assumption in this feature is somehow similar to the assumption in F_2: There will be more significant signal peaks in a noisy segment than a clean segment. A significant signal peak is defined as any sample that obtains higher amplitude than the three preceding samples and the three subsequent samples. The detection of significant signal peaks is illustrated for a noisy and a clean ECG segment in Fig. 2(a). The number of peaks in channel 1 in each window is illustrated for the entire duration of the two recordings in Fig.

TABLE I. DEMOGRAPHIC INFORMATION FOR EACH GROUP.

	Noisy Recordings	Clean Recordings
Age (mean ± std)	53.6 ± 12.9 years	50.4 ± 13.9 years
Gender	17 males, 3 females	15 males, 5 females
Recording length (mean ± std)	19.3 ± 2.0 hours	19.3 ± 1.5 hours

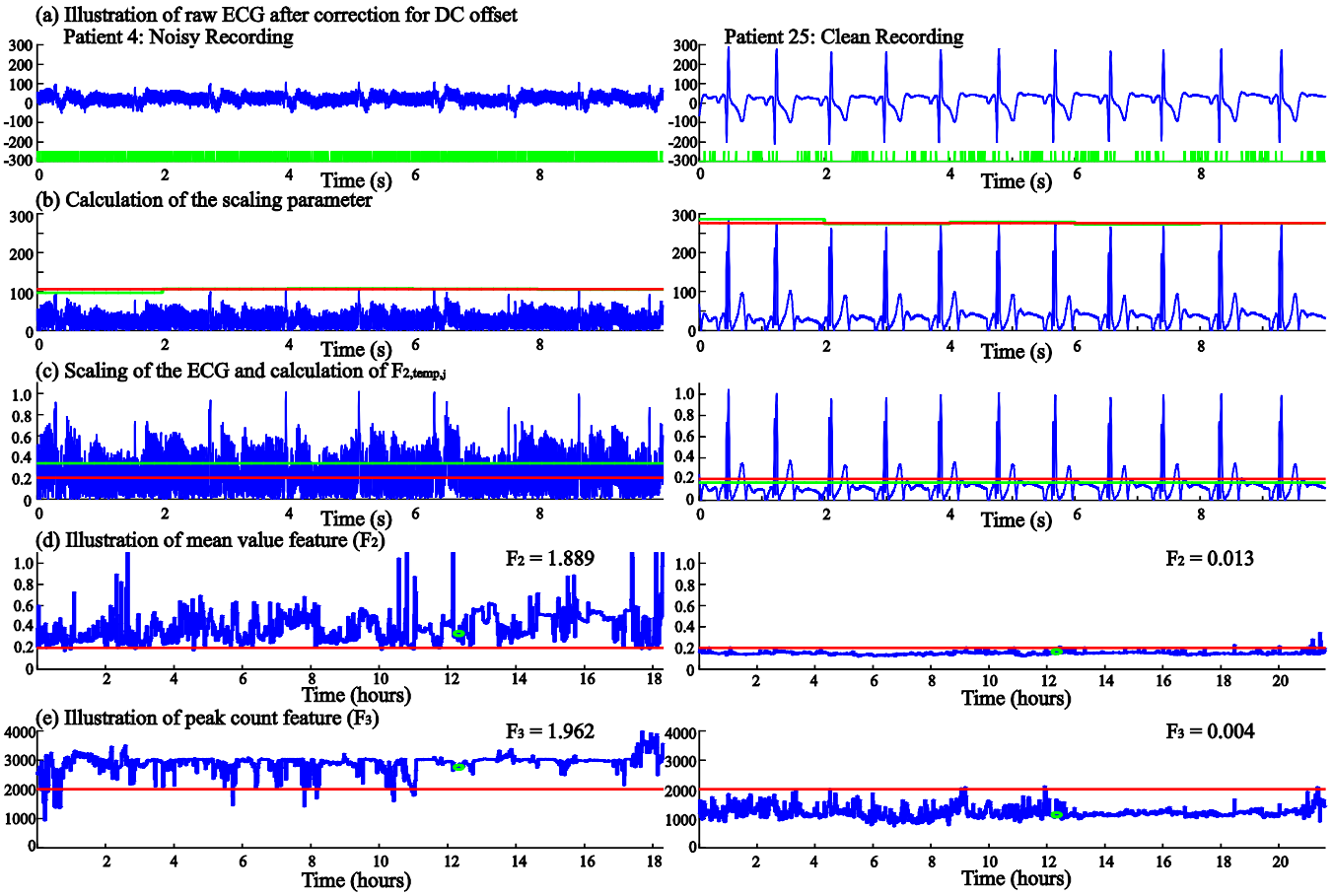

Figure 2. Illustration of feature calculation for a noisy and a clean recording (left and right panel, respectively). (a) Illustration of the raw ECG from channel 1 without any filtering. However, the DC value corresponding to 2^{11} (12 bit resolution) is removed to center the signal on 0. This raw ECG is applied for the calculation of F_2 and F_3. The green lines in the bottom indicate the samples where significant signal peaks were detected for F_3. (b) Calculation of the scaling parameter. The green lines indicate the maximum value in each two second window, and the red lines indicate the resulting scaling parameter for the current one minute window. (c) Scaling of the signal according to the scaling parameter. The green lines indicate the value of $F_{2,temp,j}$ for the current one minute window, and the red lines indicate the threshold, T_2. It is observed that $F_{2,temp,1}$ is above T_2 for the noisy segment and below for the clean segment. (d) The value of $F_{2,temp,j}$ for the entire duration of the signals. The red lines indicate T_2. (e) The number of detected signal peaks in each one minute window throughout the entire recordings. The red lines indicate T_3. The green marks in (d)-(e) indicate the position of the example illustrated in (a)-(c). The calculated feature values for each recording are also provided in (d)-(e). The value of F_1 is 0.015 and 0.000 for the noisy and clean recording, respectively.

2(e). The final feature is calculated from (4), where T_3 is a threshold, and P_j is the number of significant signal peaks in channel j. The threshold was set to 2000 by visual inspection of illustrations similar to Fig 2(e).

$$F_3 = \frac{1}{M}\sum_{j=1}^{2}\sum_{m=1}^{M} logical\{P_j(m) > T_3\} \quad (4)$$

D. Classification

For this pilot study, different types of discriminant functions were investigated for the classification task. We decided to use a simple Bayes classifier that is known to have high performance with low computational costs. A diagonal covariance matrix is applied, corresponding to assuming that the features are non-correlated. The discriminant function, d_i, is thus calculated according to (5), where $\sum_i$ is the covariance matrix of class i, "ln" is the natural logarithm operator, p_i is the prior probability of class i ($p_1 = p_2 = 0.5$), $\boldsymbol{\mu_i}$ is the mean vector of the feature vectors from class i in the training data, $\mathbf{y}$ is the feature vector to be classified ($\mathbf{y} = [F_1, F_2, F_3]$), and c is a constant equal to the natural logarithm of the determinant of the class covariance matrix, $\sum_i$.

$$d_i(\boldsymbol{y}) = -\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu_i})^T \sum_i^{-1}(\boldsymbol{y}-\boldsymbol{\mu_i}) + \ln(p_i) + c \quad (5)$$

The feature vector under classification, $\mathbf{y}$, is then classified to the class obtaining the highest value of d_i. The signal processing was conducted in MATLAB R2013b, and the classification was implemented using the build-in function "classify" with the option "diagQuadratic".

III. Results

The algorithm performance was evaluated as the sensitivity (Se=TP/(TP+FN)), specificity (Sp=TN/(TN+FP)), and accuracy (Acc=(TP+TN)/(TP+TN+FP+FN)), where TP is the number of clean recordings correctly classified as clean (true positive), TN is the number of noisy recordings correctly classified as noisy (true negative), FN is the number of clean recordings wrongly classified as noisy (false negative), and FP is the number of noisy recordings wrongly classified as clean (false positive). Due to the intermediate number of recordings in this pilot study, the performance was evaluated by a five-fold cross validation. Each fold consists of training the classifier on 32 recordings (16 from each class), and testing the performance of the obtained classifier

on the remaining eight recordings. The training and test performances for each fold, as well as the average performances are provided in Table II.

TABLE II. PERFORMANCE EVALUATION. TRAINING: *. TEST: ¤.

	Se*	Sp*	Acc*	Se¤	Sp¤	Acc¤
Fold 1	93.8%	81.3%	87.5%	100%	75%	87.5%
Fold 2	93.8%	87.5%	90.6%	75%	100%	87.5%
Fold 3	93.8%	87.5%	90.6%	100%	75%	87.5%
Fold 4	93.8%	93.8%	93.8%	100%	75%	87.5%
Fold 5	93.8%	75.0%	84.4%	100%	100%	100%
Average	93.8%	85.0%	89.4%	95%	85%	90%

IV. DISCUSSION

The proposed novel algorithm is capable of obtaining an average accuracy of 90% on the test data. This is considered a high clinical performance. It should be stated that this high performance is obtained on clinically relevant ambulatory ECG recordings acquired from real patients in their homes. This was chosen to ensure a realistic amount of abnormal heart rhythms and beat morphologies. It is, of course, very important to ensure that automatic noise classification algorithms will not classify a recording with a high number of abnormal beat morphologies as noisy. The general HR, non-disturbing baseline wandering, and different P- and T-wave morphologies might affect the values of F_2 and F_3. This was not accounted for in the calculation of the features in this pilot study. Further improvements of the algorithm might include adjustments to account for these issues. The values of T_2 and T_3 were furthermore found by visual inspection. The overall performance of the algorithm depends on the performance of these thresholds, and the overall performance might therefore be increased, if the threshold values were corrected using a proper parameter optimization method, e.g. Receiver Operator Curves (ROCs). Furthermore, some of the algorithm parameters (e.g. α and β), should be adjusted to different recording devices with different front-end specifications. However, the novel adaptive scaling of the ECG signal before calculation of F_2 ensures the possibility of a global T_2 value that is not neither patient nor device dependent. Due to the high clinical performance of the algorithm, this novel approach to quantification of noise levels in entire ECG recordings is expected to be very useful in many different applications in the future. It is extremely important to gain solid knowledge related to the benefits and drawbacks of the new technologies for long-term ambulatory ECG monitoring in different situations. Choosing the right device in each application can increase the diagnostic yield and decrease the burden on the patients and the healthcare facilities. An automatic classification of entire ECG recordings provides the possibility of an objective and fast assessment of the clinical quality of a high number of ECG recordings acquired using the different technologies. This could provide important information to answers related to the benefits and drawbacks of the new technologies. Another application scenario is related to pre-screening of recordings before the manual analysis. If a specific recording is classified as being very noisy, it might be beneficial to exclude the recording from manual analysis to increase the efficiency of the healthcare facilities. Furthermore, the algorithm is designed using very simple features that can be efficiently calculated in real-time. This could allow for future embedded implementation of the algorithm in the patch type ECG recorders. This could be imagined to provide a real-time estimate of the recorded signal quality, and allow prober actions to increase the quality if the recording quality is generally too low, or if the quality suddenly decreases during a recording. Another approach to noise estimation is using a shut-down algorithm. The shut-down approach might provide a percentage of analyzable data, but this does not necessarily directly translate to an estimate of the general quality of the entire recording: It is extremely difficult from an engineering point of view to determine the specific types, amounts, and duration of artifacts that might interfere with the clinical interpretation of a signal. The shut-down approach might be less sensitive to long periods of data with relatively poor quality that would not trigger the detection of an artifact event, but that would still impose difficulties in the interpretation of the recorded ECG signal. We therefore find it highly relevant, not only to attempt to detect the noise events, but also to provide an overall estimate of the quality of entire ECG recordings. Future work might include dividing the recording into smaller segments, and for instance disregard data based on an hour basis instead of the entire recording. It should, of course, also be stated that the algorithm performance might be further improved by exploring new features, adaptive thresholds, and more advanced classification schemes. Furthermore, the algorithm should be tested on a larger database to confirm the performance in the general population and in ECG signals with a higher variety of abnormal beat morphologies.

ACKNOWLEDGMENT

The authors wish to thank Master of Engineering Rune Paamand and the clinical staff at Glostrup Hospital (the Danish Center for Sleep Medicine, Department of Neurology, and Department of Medicine) for conducting the clinical recordings.

REFERENCES

[1] P. M. Barrett, R. Komatireddy, S. Haaser, S. Topol, J. Sheard, J. Encinas, A. J. Fought, and E. J. Topol, "Comparison of 24-hour Holter Monitoring with 14-day Novel Adhesive Patch Electrocardiographic Monitoring.," *Am. J. Med.*, vol. 127, no. 1, pp. 95.e11–7, Jan. 2014.

[2] M. P. Turakhia, D. D. Hoang, P. Zimetbaum, J. D. Miller, V. F. Froelicher, U. N. Kumar, X. Xu, F. Yang, and P. A. Heidenreich, "Diagnostic utility of a novel leadless arrhythmia monitoring device," *Am. J. Cardiol.*, vol. 112, no. 4, pp. 520–4, Aug. 2013.

[3] D. B. Saadi, I. Fauerskov, A. Osmanagic, H. M. Sheta, H. B. D. Sorensen, K. Egstrup, and K. Hoppe, "Heart rhythm analysis using ECG recorded with a novel sternum based patch technology - A pilot study," in *Cardiotechnix 2013: Proc. of the International Congress on Cardiovasular Technologies*, 2013, pp. 15–21.

[4] S. Mittal, C. Movsowitz, and J. S. Steinberg, "Ambulatory external electrocardiographic monitoring: focus on atrial fibrillation.," *J. Am. Coll. Cardiol.*, vol. 58, no. 17, pp. 1741–9, Oct. 2011.

[5] I. Silva, G. B. Moody, and L. Celi, "Improving the Quality of ECGs Collected Using Mobile PhonesThe PhysioNet / Computing in Cardiology Challenge 2011," in Computing in Cardiology 2011, 2011, pp. 273–276.

[6] "medilog Darwin Liberty Online Holter Analysis," *HASIBA Medical GmbH*. [Online]. Available: http://hasimed.com/. [Accessed: 04-Jun-2014].