An Algorithm for Assessment of Quality of ECGs Acquired via Mobile Telephones

Philip Langley, Luigi Yuri Di Marco, Susan King, David Duncan, Costanzo Di Maria, Wenfeng Duan, Marjan Bojarnejad, Dingchang Zheng, John Allen, Alan Murray

Freeman Hospital and Newcastle University, Newcastle upon Tyne, UK
DEIS, University of Bologna, Bologna, Italy

Abstract

For the application of acquiring ECGs from mobile telephones by unskilled users it would be beneficial if the mobile device could assess ECG quality and inform the user if the quality was acceptable.

Using the PhysioNet/Computing in Cardiology Challenge 2011 dataset we identified several ECG features that were commonly observed in the training set 'unacceptable' category for algorithmic development: flat baseline (FB), saturation (SA), baseline drift (BD), low amplitude (LA), high amplitude (HA) and steep slope (SS).

For the training set with each feature detection applied separately the following scores were achieved: FB 76.2%, SA 80.9%, BD 61.3%, LA 75.6%, HA 74.1% and SS 77.5%. With all features combined a score of 91.4% was achieved. For the test set the algorithm classified 181 records as unacceptable and 319 records as acceptable and the score was 85.7%.

1. Introduction

Widespread availability of mobile telephony offers the capability to acquire ECG data away from healthcare centres by novice users [1]. To acquire ECGs of diagnostic quality requires skill and is carried out by trained staff in the clinical setting. Adequate skin preparation, correct positioning of electrodes and artifact reduction are fundamental requirements [2-4]. For the application of acquiring ECGs from mobile telephones by unskilled users it would be beneficial if the mobile device could assess ECG quality and inform the user if the quality was acceptable or not, and that was the aim of this study.

2. Methods

Data were provided by the PhysioNet/Computing in Cardiology Challenge 2011[5]. Data comprised 1500 12-lead ECGs acquired on mobile telephones by users with a range of skills in ECG acquisition. Recorded ECGs were categorised as 'acceptable' or 'unacceptable' by expert annotators, although challenge participants were also encouraged to categorise the ECGs.[5] 1000 ECGs and their categories constituted a training set (Set A) and the remaining ECGs, without categories, constituted a test set (Set B). We identified by visual inspection several ECG features that were commonly observed in the 'unacceptable' category for algorithmic development: flat baseline (FB), saturation (SA), baseline drift (BD), low amplitude (LA), high amplitude (HA) and steep slope (excluding pacemaker) (SS). ECGs with any lead exhibiting these features were classed as 'unacceptable'. Example ECGs from the training set exhibiting these features are illustrated in figure 1.

2.1. Description of the algorithm

The ECG quality discrimination ('acceptable' vs. 'unacceptable') algorithm consists of the following (cascaded) detection steps:

I) Flat line (FB): constant voltage ECG excerpts of at least 1 s length are searched in all leads. If any is found the recording is classified as 'unacceptable'.

II) Saturation (SA): ECG excerpts of amplitude greater than 2 mV for more than (continuous) 200 ms are searched in all leads. If any is found the recording is classified as 'unacceptable'.

III) Baseline drift (BD): baseline is extracted from ECG by 6^{th} order Butterworth lowpass filter (Fc(3dB) = 1 Hz). If BD exceeds (excluding the first 2 s of recording) a threshold of 2.5 mV in any lead, the recording is classified as 'unacceptable'.

IV) Low amplitude (LA): after baseline subtraction, the ECG amplitude is checked on each lead. If any lead is found whose maximum amplitude falls below a threshold of 125 μV, or at least three leads are found with maximum amplitude below 175 μV, the recording is classified as 'unacceptable'.

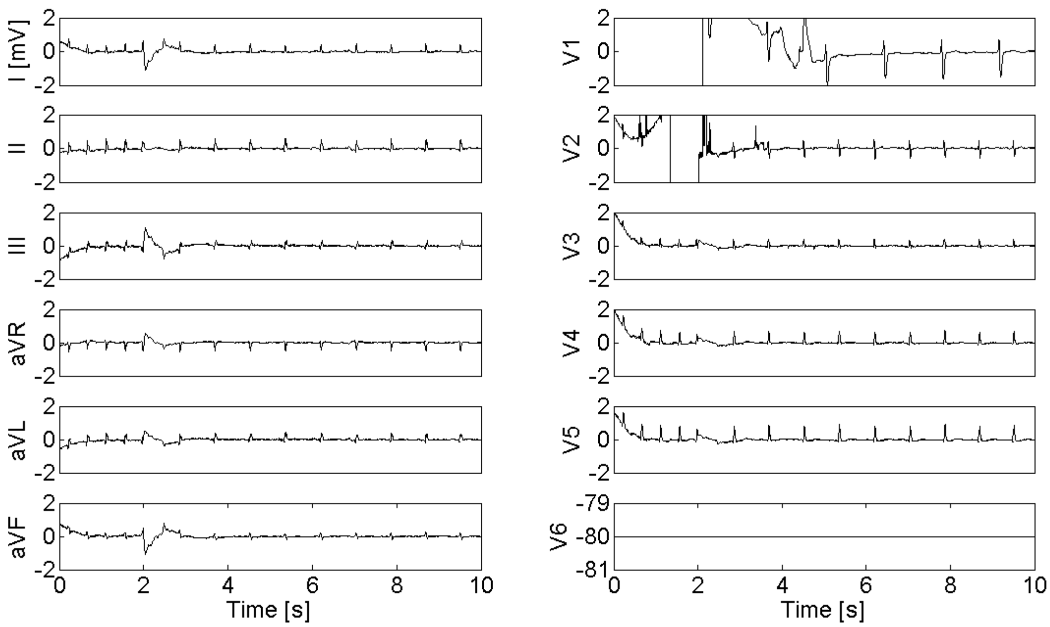

Figure 1. Example of 'unacceptable' ECG recording (# 1086219) from the training set, with BD (V1, V2), SS (V1, V2), FB (V6).

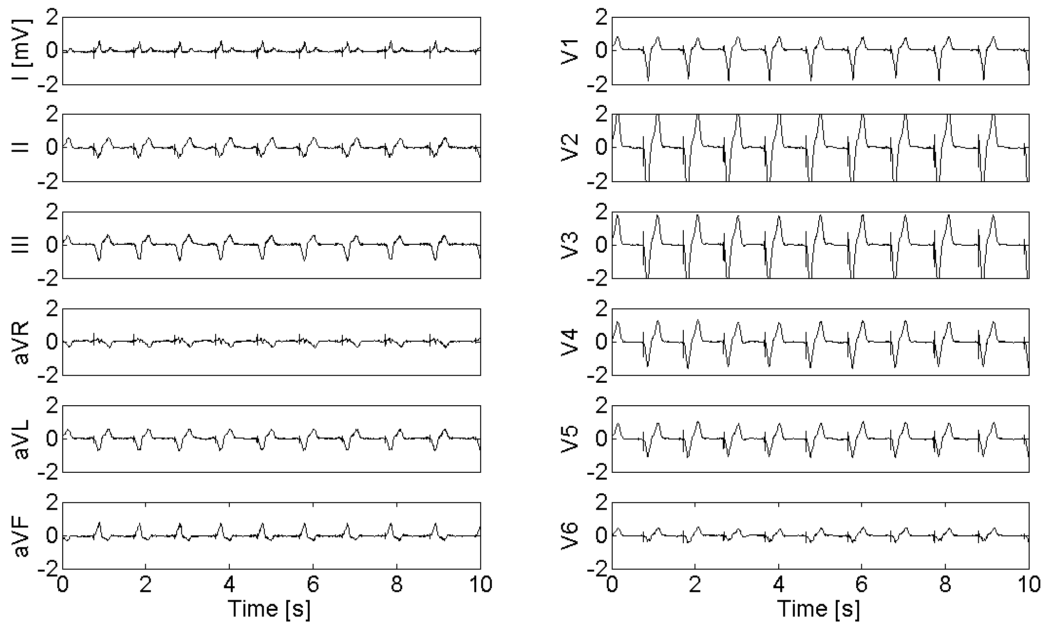

Figure 2. Example of 'acceptable' ECG recording (# 1833468) from the training set correctly classified by the algorithm. Pacemaker prominent in V2-V4.

V) High Amplitude (HA): after baseline subtraction and low-amplitude check, the ECG is checked for high amplitude on each lead. If any lead is found whose amplitude is above a threshold of 3.75 mV, the recording is classified as 'unacceptable'.

VI) Steep slope: in order to distinguish pacemaker (PM) from high-frequency noise, individual ECG leads are parsed searching for any samples t_{PM} whose slope $\Delta_{ECG}(t_{PM})$ is steeper than a threshold of 250 μV/sample. If any is found, a window W_{PM}=[t_{PM}-*80ms*, t_{PM}+*100ms*] is defined and the ECG is trimmed to a constant value throughout the window: $ECG(W_{PM}) = ECG(t_{PM}\text{-}80ms)$, and a refractory time interval of 100 ms following W_{PM} is assigned where no other t_{PM} episodes are expected. After completing this filtering stage, the filtered ECG was once again parsed on individual leads searching for spikes with slope steeper than a threshold of 1 mV/sample. If any is found, the recording is classified as 'unacceptable'.

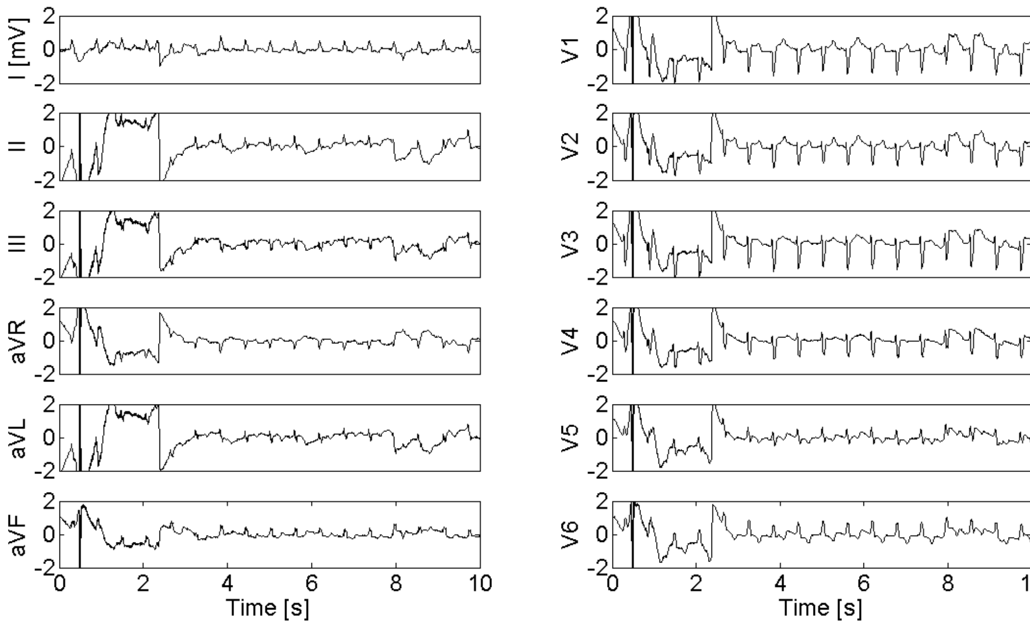

Figure 3. Example of 'acceptable' ECG recording (# 1961627) from the training set incorrectly classified as 'unacceptable' by the algorithm.

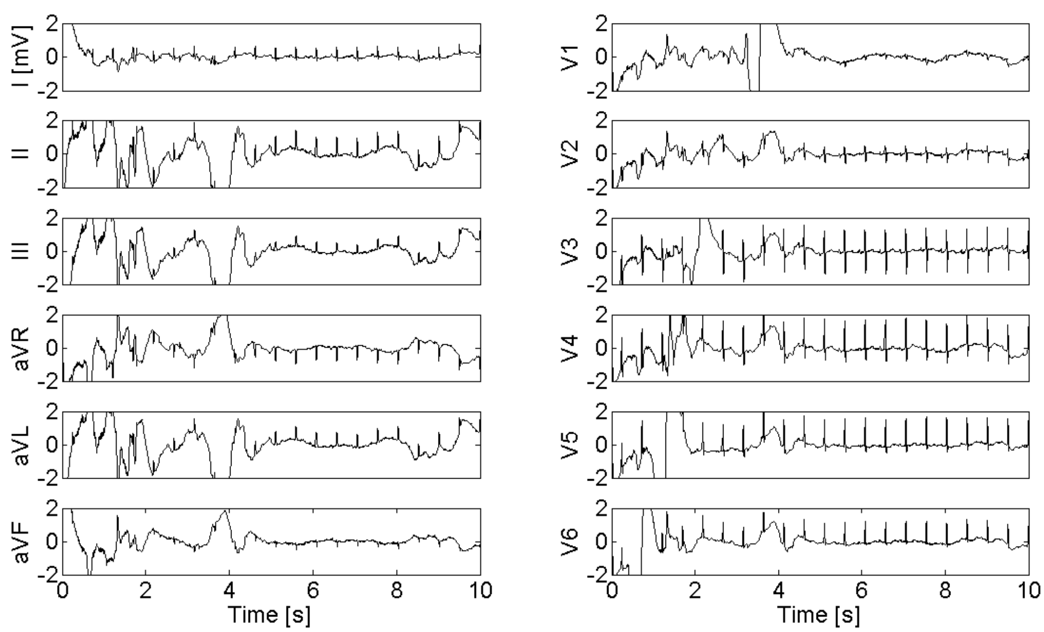

Figure 4a. Example of noisy ECG recording (# 2428645) classified as 'acceptable' in the training set.

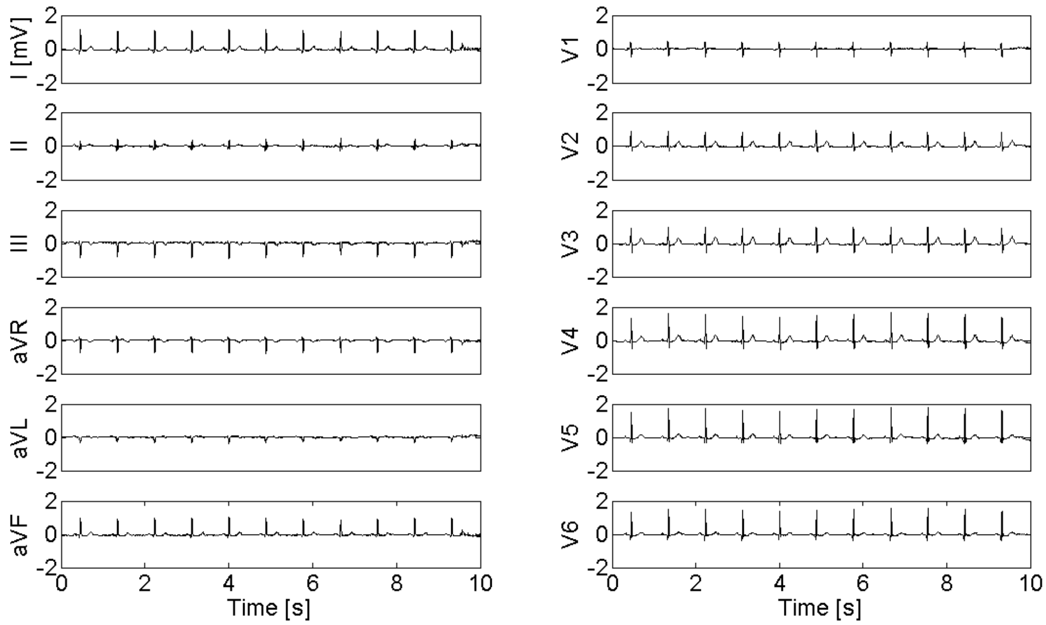

Figure 4b. Example of clean ECG recording (# 2428645) classified as 'unacceptable' in the training set.

3. Results

For the training set, with each quality discriminator applied separately, the following scores were achieved: FB 76.2%, SA 80.9%, BD 61.3%, LA 75.6%, HA 74.1% and SS 77.5%. With all features combined a score of 91.4% was achieved. Figures 2 and 3 provide examples of ECGs from the training set which our algorithm classified correctly and incorrectly (for ECG classifications on 1st May 2011).

For the test set the algorithm classified 181 records as unacceptable and 319 records as acceptable and the score was 85.7% (for ECG classifications on May 1st 2011).

4. Discussion

Because this algorithm is designed for the mobile phone platform, only linear time-invariant (LTI) methods were used currently to keep the computational complexity and power consumption at a relatively low level. In the future, frequency domain analysis methods could be added, which are effective to detect some other kinds of noise, such the powerline coupling (50/60 Hz), EMG noise and low frequency baseline modulation [6].

Some records from the training set classified by experts as 'acceptable' were characterized by remarkably noisy signals as shown in figure 4a. On the other hand, some records classified as 'unacceptable' apparently had high quality signals, as shown in figure 4b. The algorithm score was penalized by such cases.

References

[1] Oresko JJ, Duschl H, Cheng AC. A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing. IEEE Trans Inf Technol Biomed 2010; 14: 734-40.

[2] Davies A. Recognizing and reducing interference on 12-lead electrocardiograms. Br J Nurs 2007; 16: 800-4.

[3] Drew BJ. Pitfalls and artifacts in electrocardiography. Cardiol Clin 2006; 24: 309-15.

[4] García-Niebla J, Llontop-García P, Valle-Racero JI, Serra-Autonell G, Batchvarov VN, de Luna AB. Technical mistakes during the acquisition of the electrocardiogram. Ann Noninvasive Electrocardiol 2009; 14: 389-403.

[5] Moody G. Computing in cardiology challenge 2011. Computing in Cardiology 2011.

[6] Allen J, Murray A. Assessing ECG signal quality on a coronary care unit. Physiol Meas 1996; 17: 249-58.

Address for correspondence.

Philip Langley
Medical Physics Dept
Freeman Hospital
Newcastle upon Tyne
UK
philip.langley@ncl.ac.uk