A database of simultaneously recorded ECG signals with and without EMG noise

Vladimir Atanasoski, Jovana Petrovic, Lana Popović Maneski, Marjan Miletić, Miloš Babić, Aleksandra Nikolić, Dorin Panescu, *Fellow, IEEE,* and Marija D. Ivanović

Abstract— Goal: **Noise on recorded electrocardiographic (ECG) signals may affect their clinical interpretation. Electromyographic (EMG) noise spectrally coincides with the QRS complex, which makes its removal particularly challenging. The problem of evaluating the noise-removal techniques has commonly been approached by algorithm testing on the contaminated ECG signals constructed *ad hoc* as an additive mixture of a noise-free ECG signal and noise. Consequently, there is an absence of a unique/standard database for testing and comparing different denoising methods. We present a SimEMG database recorded by a novel acquisition method that allows for direct recording of the genuine EMG-noise-free and -contaminated ECG signals. The database is available as open source.**

Index Terms— ECG acquisition, EMG noise, denoising

Impact Statement— **SimEMG is the first dataset with simultaneously recorded electrocardiograms with and without electromyographic noise. It provides unique set for direct assessment and comparison of ECG denoising and filtering methods.**

I. INTRODUCTION

WHEN performed outside the clinical settings, ECG measurements are highly susceptible to noise. The rapidly increasing demand for mobile ECG devices has renewed the interest in overcoming this challenge. Common noise sources are: baseline wander (BLW), power-line interference (PLI), motion artifacts, and electromyographic (EMG) noise [1]. The EMG noise is broadband (> 10 Hz) and particularly challenging to eliminate due to its significant spectral overlap with the QRS complex – the principal information carrier in ECG signal. The EMG noise origin in involuntary muscle movements does not allow for an obvious mitigation strategy. Hence, the EMG noise must be eliminated

Vladimir Atanasoski, Jovana Petrovic, Marjan Miletić and Marija D. Ivanović are with Vinca Institute of Nuclear Sciences, Mike Petrovica Alasa 12-14, Belgrade, Serbia and HeartBeam, Inc. 2118 Walsh Avenue, Suite 210, Santa Clara, CA, USA (Correspondence e-mail: vladimir.atanasoski@vin.bg.ac.rs). Lana Popović Maneski is with Department of Institute of Technical Sciences of the SASA, Knez Mihailova 35/IV, Belgrade, Serbia. Miloš Babic and Aleksandra Nikolić are with the Institute for Cardiovascular Diseases Dedinje, Serbia, Heroja Milana Tepica 1, Belgrade, Serbia. Dorin Panescu is with BIOTRONIK CRC EP, California, USA.

by signal processing, which is particularly true for the measurements by handheld ECG devices, which imply the engagement of fingers [2], [3].

A number of EMG-noise removal techniques have been developed, including various time and frequency filtering [4], [5], and deep learning [6], [7] methods. However, evaluation and comparison of these techniques remain a problem due to the nonexistence of a suitable evaluation data set. This problem has commonly been approached by using the contaminated ECG signals constructed as an additive mixture of noise-free ECG signals and noise. The noise-free ECGs are typically taken from publically available databases by selecting segments without considerable noise levels or obtained artificially by ECG signal generators. The noise is typically taken from the MIT noise stress test (MIT-NST) database, which comprises a distinct original EMG signal and other noise sources [8,] [9] or derived by nonlinear filtering of white Gaussian noise to spectrally match the EMG noise [5], [10]. Signal selection and construction may bias the evaluation and comparison of denoising algorithms. However, reports on the evaluation and comparison of denoising algorithms vary widely in signal selection and construction. For example, reported are EMG-contaminated ECG signals obtained from the CSE database with added CGN noise [5], from the MIT-BIH Normal Sinus Rhythm database with CGN noise [10]; from the MIT-BIH database with the noise procured from the MIT-NST database [11], synthetized ECG signals with added CGN noise [12], etc. Therefore, it is difficult to establish an optimum denoising technique and evaluate its performance.

Here, we propose and implement a new acquisition method, SimEMG, that allows simultaneous recording of EMG-noise-contaminated and free ECG signals. The method relies on the reference measurement performed with the ECG electrodes placed on the upper arm, which is known to be much less affected by EMG noise than the hands. We apply the method to create an open SimEMG database, the first database of genuine noisy and EMG-noise-free signals suitable for evaluating and comparing denoising algorithms.

II. DATA COLLECTION PROCEDURES

This work was approved by the Human Research Ethics Committee of the Institute for Cardiovascular Diseases Dedinje, Serbia, with the ethical standards of the institutional and/or national research (approval number 4474).

A. Acquisition method

The method assumes that the potential on every point along the arm is constant when muscles are relaxed [13] and that the EMG noise is generated locally in the hands. Then, the potential difference between electrodes placed on the hands includes the EMG noise, while the potential difference between electrodes placed on the shoulders distant from the noise source renders the reference signal without noise (Fig. 1).
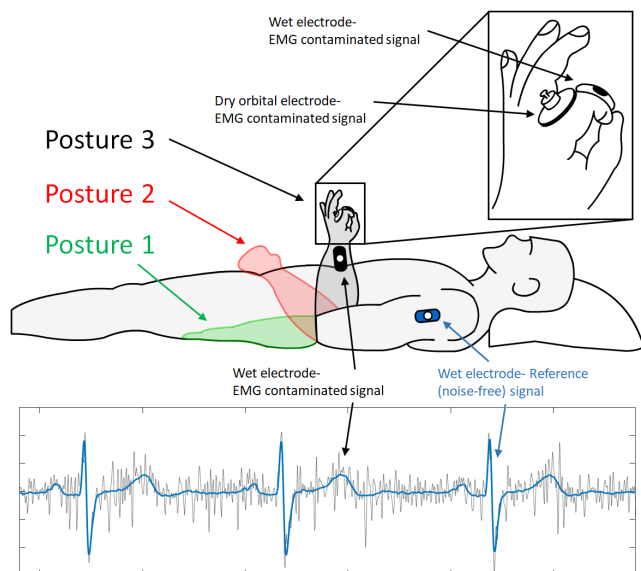

Fig. 1. SimECG acquisition method. Noisy signals (grey line) were recorded in three different forearm Postures 1-3. Noise-free signals (blue line) were obtained from the upper arm electrode. The same postures were assumed by both arms.

The standard clinical ECG includes recordings from 12 leads (3 limb leads, 3 augmented limb leads, and 6 precordial leads) obtained by placing 10 electrodes at predefined positions on the body surface (2 on arms, 2 on legs, and 6 on chest). All precordial electrodes (V1-V6) are referenced to the Wilson central terminal. Hence, any pair of precordial electrodes can be used to measure the voltage difference between any two points on the body. For example, a pair of precordial electrodes placed in the positions of lead I electrodes will give the potential difference equivalent to the lead I. We exploit this to define SimEMG electrode positioning in the following manner:

- The electrodes for recording the limb leads are placed in the standard manner.
- A pair of precordial electrodes (V1 and V2) is placed on the shoulders at the point of attachment of the deltoid muscle (shoulder muscle) to the humerus (upper arm) to obtain the reference noise-free signal.
- The other two pairs of precordial electrodes are placed in the regions of intermediate (V3 and V4) and proximal (V5 and V6) phalanges on the upper side of the left and right hands to record signals contaminated by EMG noise.

Note that here, the term 'precordial' refers to the standard ECG hardware and does not describe the placement of

electrodes in SimEMG configuration.

Each SimEMG measurement generates 4 single-lead signals obtained by subtracting the potentials within each electrode pair: a reference signal from the shoulders and three signals contaminated with noise (one from the standard limb leads and two from the potential differences between fingers of the left and right hand). The EMG noise is introduced by activating hand muscles either by pressing two fingers of one hand against each other or pressing the object that causes resistance. An example of recorded signals is shown in Fig. 2.
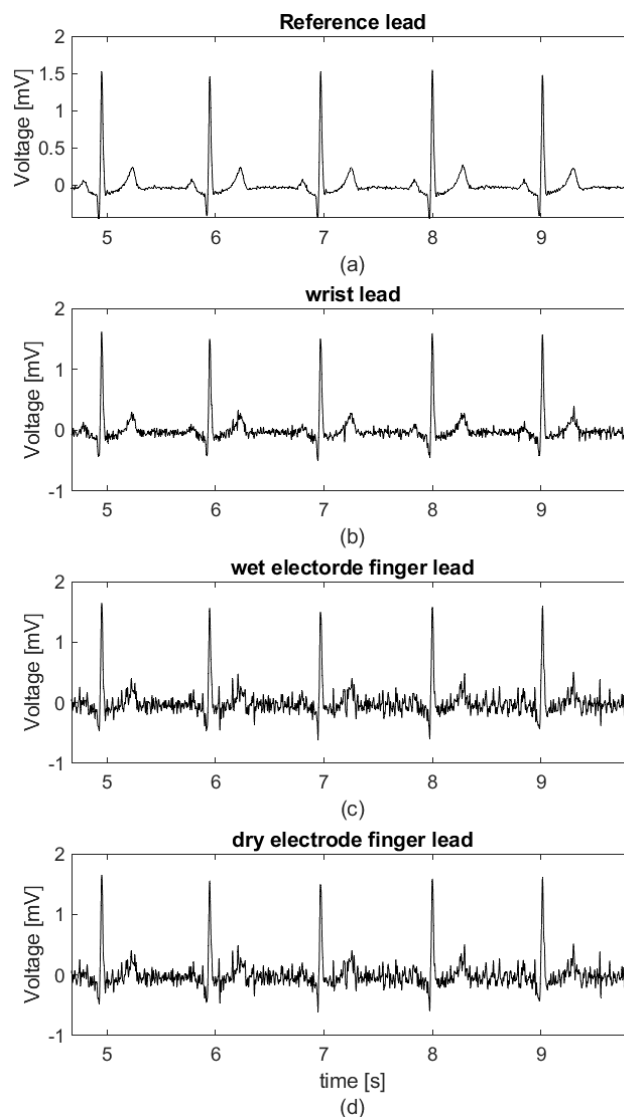

Fig. 2. A typical set of signals from a single measurement. a) Reference lead, b) wrist lead, c) wet electrode finger lead, d) dry electrode finger lead.

Here, ECG was recorded by a custom-made 12-lead ECG device based on a 24-bit Texas Instruments A/D converter. The recordings were 30 seconds long with a sampling rate of 500 Hz. The amplitude resolution was 200 samples per mV. The reference, limb, and one finger-lead were recorded by the standard Ag/AgCl electrodes with gel (Ag/AgCl gel) [14] commonly used in clinics. Despite the excellent performance of the Ag/AgCl gel electrodes in recording ECG signals, they

are not suitable for mobile ECG devices. It has been shown that in mobile ECG monitoring, the orbital electrodes have better performance [16]. Hence, the second finger lead was recorded by a pair of orbital electrodes (Orbital Research Inc., Cleveland, OH, USA) [15] fixed to the fingers with a medical adhesive tape.

B. Recording protocol

Upon signing the written consent, 15 healthy volunteers participated in the study, 6 males and 9 females, aged 44.5±14.6. Measurements were performed with the subjects in the supine position and three different arm postures shown in Fig. 1 and explained below:
- the arms placed along the body (posture 1),
- the forearms leaned on the hips, approximately at the angle of 45 degrees relative to the bed (posture 2),
- the forearms point upright, with elbows supported on the bed next to the body (posture 3).

Subjects were holding the clips in their hands and were instructed to press them by thumb and index finger to generate the EMG noise. They were instructed to stay relaxed as much as possible in their upper arms so that the reference signal had a minor EMG component detected. We collected three recordings per arm posture, which resulted in 9 overall recordings per subject.

C. Signal processing

In order to obtain signals contaminated only with EMG noise, we applied the following signal processing:
- Baseline wander was removed by a 5th–order high-pass Butterworth filter with a cutoff frequency of 1 Hz;
- Power-line interference was removed by applying the 2nd order IIR notch filter with the central frequency of 50 Hz;
- Low-pass 2nd–order Butterworth filter with a cutoff frequency of 100 Hz was utilized for removal of the high-frequency components (beyond the spectrum of the QRS complex).

The postprocessing was performed in MATLAB, MathWorks Inc.

III. VALIDATION PROCEDURES

All recordings were approved by a cardiologist, who inspected and validated the ECG electrode positions. One single-lead recording is rejected due to electrode misplacing.

A number of the recorded reference signals contained a small residual EMG noise. To retain as clean reference signals as possible, we established a rejection criterion based on the estimated noise level. From a signal filtered by a high-pass filter with a cutoff frequency of 25 Hz, we first excluded all the samples contained in the interval from R-60ms to R+60ms (QRS complex) and then calculated the root-mean-square error (RMSE) on the remaining samples to use it as a representation of the noise content. All reference signals with RMSE > 0.011 mV were rejected, whereby the RMSE threshold was set empirically. This resulted in the exclusion of 89 out of 126 recordings.

We set two successful reference recordings taken at complete rest in the lying position as an inclusion criterion for subjects. If a subject did not meet this criterion, all recordings from that subject were excluded from the study. This resulted in the exclusion of one subject.

One subject could not make a satisfactory reference signal when introducing EMG noise in both hands. To reduce the residual noise during recording, we suggested engaging only one hand in generating the EMG noise component. This maneuver resulted in a reference signal with the RMSE below the threshold.

IV. DATASET DESCRIPTION

The recordings are stored and published in a repository in Mendeley [17]. SimEMG contains 147 signals in total, 37 out of which are noise-free and 110 noise-contaminated single-lead recordings. They are generated from 14 subjects (9 females and 5 males aged 40±13). Each recording is 30 seconds long.

Recordings are saved in ASCII format in the files named 'PatientNumber_RecordingNumber_ElectrodePosition.mat', where PatientNumber counts patients and assumes values P1, P2, … P14, RecordingNumber counts recordings per patient, and ElectrodePosition names the recording lead: 'lead I' for the wrist lead recorded with Ag/AgCl electrodes, 'Ag-AgCl' for the finger lead recorded by Ag/AgCl electrodes, and 'ORB' for the finger lead recorded with orbital electrodes. For example, 'P3_2_Ag-AgCl.txt' is a finger lead recorded with gel electrodes in the second measurement on patient 3.

To quantify the level of noise, we evaluated the signal-to-noise ratio, defined as the ratio between the recorded reference signal and EMG noise, on all signals. Here, the noise was defined as the difference between a noise-contaminated and the reference signal. The average SNR of the noise-contaminated signals was 8.53±5.5 dB. Most signals obtained from fingers had SNR < 8 dB, while only 7 recordings had SNR > 16 dB, indicating a high overall noise level (Fig. 3).
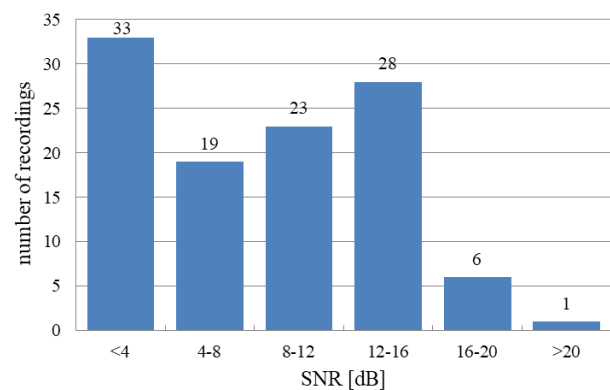

Fig. 3. Number of recordings at different noise levels

V. USAGE NOTES

The SimEMG database is significant as it enables direct

evaluation and comparison of the methods for EMG noise removal from ECG signals. This is particularly important in mobile hand-held devices in which the involuntary muscle movement in the hands generates large EMG noise. All algorithms developed for this purpose, including ensemble averaging (EA), wavelet transformation, adaptive filtering, independent component analysis, and model-based filtering [4], [5], [10], [12], can be evaluated and compared using SimEMG. Moreover, SimEMG can also serve as a test set for denoising methods based on deep learning, such as autoencoders [6] or U-net-like networks [7]. Its use as a learning set is currently limited by the number of recordings. However, the proposed acquisition method and protocol can be used to generate larger sets.

Although insignificant for diagnostic purposes, the small amount of residual EMG noise can influence the comparison of different filtering methods. Our future work will focus on addressing this issue. One possibility would be to record a multi-lead ECG signal on the torso in the vicinity of the arms and use it to reconstruct a reference signal. This allows for the recording of the noise-free signals along with the signals with EMG noise, BLW, and motion artifact, thus enabling full comparison of denoising methods on real ECG signals.

VI. Code Availability

The MATLAB (MathWorks, Inc.) code used for data processing is provided here

https://drive.google.com/file/d/128EFArKpYfFMkcIrvzadP8h EKlk8shKt/view?usp=share_link

Author contributions

V.A. and L.P.M. have conceived the idea and constructed the measurement protocol. V.A., L.P. M, M.M., M.B. and A.N. have performed and validated measurements. V.A. has processed the signals. V.A., M.B. and A.N. were responsible for data curation, V.A., J.P., D.P. and M.D.I. have analysed and validated the results. V.A., J.P., M.B., D.P. and M.D.I. have written the paper. J.P. and A.N. have secured funding and managed the project.

Statement of the conflict of interest

V.A., J.P., M.M., L.P.M. and M.D.I. declare that the SimEMG method was conceived and database constructed for the purposes of evaluation of the EMG noise filters developed by Heart Beam, Inc. The method and database are published as open source and as such fully available to others.

References

[1] L. Sornmo, P. Laguna, "Noise and Artifacts," in *Bioelectrical signal processign in cardiac and neurological applications*, 1st ed. Burlington, AP, 2005, pp. 441.

[2] K. H. C. Li et al. (2019, Feb.). The Current State of Mobile Phone Apps for Monitoring Heart Rate, Heart Rate Variability, and Atrial Fibrillation: Narrative Review. *JMIR Mhealth Uhealth.* [Online]. *7(2)*. pp. e11606. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6396075/

[3] A. D. William, et al. (2018 Oct.). Assessing the accuracy of an automated atrial fibrillation detection algorithm using smartphone technology: The iREAD Study. *Heart Rhythm.* [Online]. *15(10)*, pp. 1561-1565. Available: https://www.heartrhythmjournal.com/article/S1547-5271(18)30661-1/fulltext

[4] E. Ebrahimzadeh et al. (2015) "ECG signals noise removal: selection and optimization of the best adaptive filtering algorithm based on various algorithms comparison. *Biomed. Eng. - Appl. Basis Commun.* [Online]. *27(4)*. pp. 1550038. Available https://www.worldscientific.com/doi/10.4015/S1016237215500386

[5] L. Smital et al.(2013, Feb.). Adaptive Wavelet Wiener Filtering of ECG Signals. *IEEE. Trans. Biomed. Eng.* [Online]. 60(2). pp. 437-445. Available: https://ieeexplore.ieee.org/document/6357230

[6] H. T. Chiang et al. (2019, Apr.). Noise Reduction in ECG signals Using Fully Convolutional Denoising Autoencoders. *IEEE Access.* [Online]. *7*. pp. 60806-60813. Available: https://ieeexplore.ieee.org/document/8693790

[7] A. R.-Meymandi, A. Ghaffari (2022, Jan.) A deep learning-based framework For ECG signal denoising based on stacked cardiac cycle tensor. *Biomed Signal Process Control* [Online]. *71(B)*. Available: https://www.sciencedirect.com/science/article/abs/pii/S1746809421008727

[8] A. L. Goldberger et al. (2000, June). PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circ.* [Online]. *101(23)*. pp. 215-220. Available: https://www.ahajournals.org/doi/10.1161/01.CIR.101.23.e215

[9] G.B. Moody et al. (1984). A noise stress test for arrhythmia detectors. Computers in Cardiology. 11:381-384.

[10] R. Sameni et al. (2007, Dec.). A Nonlinear Bayesian Filtering Framework for ECG Denoising. *IEEE Trans Biomed Eng.* [Online]. *54(12)*. pp.2172-2185. Available: https://ieeexplore.ieee.org/document/4360034

[11] The MIT-BIH Normal Sinus Rhythm Database. [Online]. Available: http://www.physionet.org/physiobank/database/nsrdb/

[12] P. E. McSharry and G. D. Clifford (2004, May). A comparison of nonlinear noise reduction and independent component analysis using a realistic dynamical model of the electrocardiogram. Presented at Second International Symposium on Fluctuations and Noise. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/5467/1/A-comparison-of-nonlinear-noise-reduction-and-independent-component-analysis/10.1117/12.548726.full

[13] F.N. Wilson (1930, June). The distribution of the potential differences produced by the heart beat within the body and at its surface. *Am. Heart J.* [Online]. *5(3)*. pp. 599-616. Available: https://www.sciencedirect.com/science/article/pii/S0002870330902542

[14] http://www.ceracarta.it/prodotti/nm-2844-rfs/

[15] https://www.orbitalresearch.com/medical-devices

[16] L. P. Maneski et al. (2020, Aug.). Properties of different types of dry electrodes for wearable smart monitoring devices. *Biomed Tech (Berl).* [Online]. *65(4)*. pp.405-415. Available: https://www.degruyter.com/document/doi/10.1515/bmt-2019-0167/html

[17] V.Atanasoski et al. (2022, Dec.). SimEMG database- simultaneous recordings of noise-free and noise-contaminated ECG signals. Mendeley Data, V1. Available: https://data.mendeley.com/datasets/yx5pb66hwz