Check Your Biosignals Here: A new dataset for off-the-person ECG biometrics

CrossMark

Hugo Plácido da Silva [a,*], *André Lourenço* [a,b], *Ana Fred* [a], *Nuno Raposo* [c], *Marta Aires-de-Sousa* [c]

[a] *Instituto de Telecomunicações, Instituto Superior Técnico, 1049-001 Lisboa, Portugal*
[b] *Instituto Superior de Engenharia de Lisboa, 1959-007 Lisboa, Portugal*
[c] *Escola Superior de Saúde, Cruz Vermelha Portuguesa, 1300-125 Lisboa, Portugal*

ARTICLE INFO

ABSTRACT

The Check Your Biosignals Here initiative (CYBHi) was developed as a way of creating a dataset and consistently repeatable acquisition framework, to further extend research in electrocardiographic (ECG) biometrics. In particular, our work targets the novel trend towards off-the-person data acquisition, which opens a broad new set of challenges and opportunities both for research and industry. While datasets with ECG signals collected using medical grade equipment at the chest can be easily found, for off-the-person ECG data the solution is generally for each team to collect their own corpus at considerable expense of resources. In this paper we describe the context, experimental considerations, methods, and preliminary findings of two public datasets created by our team, one for short-term and another for long-term assessment, with ECG data collected at the hand palms and fingers.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

A research field that has seen a recent trend towards the use of biosignals is biometrics, which focuses on identity recognition based on physiological or behavioral properties of an individual [1]. Electrocardiographic (ECG) data is one of the novel biometric traits where a growing interest is evident within the reference literature [2–6]. A particularly important aspect, that is transversal to all the work done to date, is the access to large datasets to evaluate the robustness of the devised methods across research teams. Over the years, several initiatives have greatly contributed to mitigate this problem, mostly by providing repositories for data collected in clinical and laboratory settings [7].

Our research group recently pivoted towards a different approach entitled "off-the-person" [8–11]. A recurrent problem faced not only by us, but also by other researchers throughout the world, is the access to large datasets, in particular collected in a repeatable manner, and that can be publicly accessible. This concern is especially important when dealing with ECG acquired in a non-traditional fashion, with minimally constrained or unconstrained scenarios, and that can account for changes in the environmental conditions, aging, and other factors surrounding the subjects.

The Check Your Biosignals Here initiative (CYBHi) was motivated by the lack of datasets of ECG signals collected at the hand palms and fingers (i.e. off-the-person). What is generally found in the state-of-the-art is that each team collects

Table 1 – Summary of datasets found in the ECG biometrics literature.

Name	MIT-BIH [7]	AHA [7]	PTB [7]	DRIVEDB [7]	CYBHi	Chan [3]	Zhao [6]	Odinaka [4]
Number of subjects	48	155	290	17	125+	50	22	265
Sampling rate	360 Hz	250 Hz	10 kHz	496 Hz	1 kHz	1 kHz	200 Hz	1 kHz
Resolution	11-bit	12-bit	16-bit	14-bit	12-bit	12-bit	n.a.	n.a.
Channels	2	2	14	1	2	1	1	1
Contact points	n.a.	n.a.	n.a.	3	2	2	3	n.a.
Electrodes	Gel	Gel	Gel	Gel	Dry Ag/AgCl + electrolycras	Dry Ag/AgCl	Gel	Gel
Placement	Chest	Chest	Chest	Chest	Hand palms + fingers	Fingers	Fingers	Chest
Publicly available	Yes	Yes	Yes	Yes	Yes	No	No	No

its own corpora, consisting of only a few dozens of individuals or less, at great financial and time expense, which is then only used in the context of their own research, and often lacks several relevant details regarding acquisition and subject conditions. As such, we devised a data acquisition framework and experimental setup, for large scale data collection from a large group of subjects through an easily repeatable and efficient procedure.

The rest of the document is organized as follows: Section 2 outlines the current landscape in terms of ECG datasets; Section 3 describes the experimental setup, highlighting the main technical challenges and options to overcome them; Sections 4 and 5 describe the adopted methods and procedures for data acquisition; Section 6 presents the findings resulting from research done to date with CYBHi; Section 7 describes the resources made available as a result of our work; and finally, Sections 8 and 9 provide a short discussion and insight on future work resulting from the CYBHi.

2. Background

Recognizing the need and usefulness of centralized datasets that can be used as a common reference for researchers worldwide, several initiatives have been contributing with resources to mitigate this. Physionet [7] is currently one of the main forums for dissemination and exchange of biomedical signals in general. Most publicly available datasets are currently centralized on the PhysioBank repository, and contain multiple parameters from healthy and pathological conditions (cardiorespiratory, neural and others). However, in what concerns ECG datasets, public resources found to date only contain signals collected at the chest with clinical grade equipment. Table 1 highlights the specifications of datasets commonly found in ECG biometrics literature.

Electrocardiographic signal acquisition methods can be classified according to their intrusiveness as follows [8]: (a) *In-the-person:* referring to implantable devices such as artificial cardiac pacemakers and implantable loop recorders; (b) *On-the-person:* referring to devices that need to be attached to the body of the subject, generally requiring conductive paste or gel; and (c) *Off-the-person:* referring to devices that are integrated in objects or surfaces with which the subjects interact with (e.g. a computer keyboard), and do not require any special preparation of the subject.

Research so far has shown broad evidence regarding the applicability of ECG signals collected in an on-the-person approach for biometric purposes (see [2,4,5] and references therein). Nonetheless, this approach is highly intrusive for the

subjects, thus limiting the potential industrial applications of ECG-based biometrics. More recently, researchers have been pivoting towards the off-the-person approaches [3,6,10–13], in an effort to overcome the intrusiveness problems widely criticized in ECG biometrics. Recent work from our team has provided experimental evidence that the signals obtained through the off-the-person approach using dry electrodes, present virtually the same morphology as those obtained at the Lead I derivation of a standard 12-lead medical-grade ECG system [8,14].

Still, to further progress on the study of this particular research topic, no public datasets currently exist (Table 1). Furthermore, in the non-public off-the-person datasets, the amount of information provided regarding the materials and methods often lacks several relevant details. In this paper, we seek to contribute to fill this gap, by providing a detailed description of our experimental setup for data acquisition, with the purpose of creating a large and extendable corpora of data, targeting the specific class of off-the-person and minimally constrained signal acquisition scenarios [8]. Our goal is to set the ground for a scalable family of datasets that enables researchers to benchmark their algorithms taking into account factors such as the permanence of the biosignals throughout time, emotional arousal, among others. Initial acquisition sessions targeted the collection of ECG signals in highly non-intrusive scenarios, and our dataset is publicly available for the community at large.

3. Design considerations

3.1. Overview

Given the motivation of our study towards the off-the-person approach, we proposed a setup for ECG data acquisition at the hand palms with dry Ag/AgCl electrodes and at the fingers with Electrolycras, to assess the biometric potential of signals collected at these anatomic regions and compare the performance of both materials. Moreover, our protocol includes both neutral tasks and emotional elicitation tasks [15,16]. The later were introduced as a way of inducing intra-subject variability, which is an important aspect to take into account when benchmarking ECG-based biometric algorithms. For this purpose, electrodermal activity (EDA) data was simultaneously acquired as a way of providing ground truth information about the arousal state of the subject, and which can be used in correlation with the ECG data.

The inclusion of emotion eliciting stimuli was motivated by the fact that authors in the field of ECG-biometrics are

(a) Volunteer station.　　(b) Expert station.

Fig. 1 – Overall workbench setting for data acquisition.

(a) edaPLUX sensor.　　(b) electrodes.

Fig. 2 – Electrodermal activity sensor and electrodes.

Fig. 3 – Electrocardiography sensors and electrodes arrangement; on the top we can see the electrolycra strips, and on the bottom the dry Ag/AgCl electrodes.

starting to explore the relation between emotional arousal and changes to the ECG heartbeat waveform [17,18]. In [17], visual stimuli were used to induce positive and negative emotions for valence and passive arousal performance evaluation, and a video game was used to attempt to induce active mental stress. In [18] the ECG was acquired during human–computer interaction on different cognitive activities. Both studies report the existence of slight morphological changes to the ECG heartbeat waveform associated with the presence of different emotional/affective states.

Fig. 1, shows the overall workbench setting, comprised of a device to display multimedia content to the subject, the set of sensors used in the experiments, and the biosignal acquisition equipment, together with the base stations for real-time data recording. The arrangement was such that the subject sat on one side of the table, and the experts on the other side. All participants were submitted to an initial briefing with the support of an informed consent document (Fig. 9). This form was read to the subject by the coordinator of the experiments, and clearly stated the materials involved in the experiments,

the purpose of our research, and the different steps that would be performed.

To acknowledge the participation as a volunteer and grant the anonymous use of the collected data, subjects were required to sign the document. Participants were also inquired about the possibility of participating in future data acquisition sessions, and primary contact information was collected for this purpose.[1] Demographic information was also requested for statistics and sample characterization purposes,

---

[1] It is important to further reinforce that contact information is stored separately from the data and that the database is anonymized.

Table 2 – Video sequences characterization.			
Seq.	Type	Timeframe	Content
A1	Low	0–55 s	Sony Bravia paint ad
	Arousal	55–60 s	Monkey falls of tree
A2	High	0–55 s	Sony Bravia bouncy balls ad
	Arousal	55–60 s	Final part of REC movie trailer

Table 3 – Specifications of the ECG and EDA sensors.		
	ECG	EDA
Filtering	1–30 Hz bandpass	3 Hz lowpass
Input Impedance	>1 MΩ	>1 TΩ
CMRR	110 dB	110 dB
Gain	1000	–

and a space for additional notes or comments regarding the experiments was also added. It is important to highlight that all the data publicly available in the CYBHi dataset is completely anonymized to preserve the privacy of the users.

3.2. Computer task

In the design of the experiment, we introduced a set of film clips that would serve as an activity for the user to perform while the signal acquisition process was taking place, but that would also elicit non-specific affective responses. In order to potentiate noticeable reactions in the subjects we manipulated the movies to introduce a stimuli. Two MPEG4 video sequences were prepared, with a 1280×720 pixel resolution and total duration of 1 min; a background soundtrack was also maintained for greater immersion and stimulatory potential. In these sequences, the last 5 s were adjusted in order to present a triggering stimuli.

To allow synchronization with the biosignal acquisition system, and also to mark the beginning of the video and the stimuli presentation, a white square with 120×120 pixels was inserted on the top left corner of the screen, which appeared for 2 s in the beginning of the video and in the moment where the triggering stimuli is presented. An Apple iPad 2 tablet device was used to display and reproduce the multimedia content; the device was placed on a vertical stand in front of the person (Fig. 12(a)), and a set of headphones were used, to provide greater focus on the reproduced content and maximize the impact of the prepared stimuli (Fig. 12(b)).

Table 2 describes the content used to produce each stimuli. Sequence A1 was designed as a more amusing video, to stimulate low arousal; it started with an excerpt of an ad with a fairly intense audio soundtrack and surprising moments,[2] which switched to a short funny video near the end.[3] Sequence A2 was designed as a more intense video to stimulate high arousal; it started with a peaceful sequence extracted from an ad,[4] and switched to an excerpt of a horror movie trailer.[5] Fig. 7 shows sample frames from both affective elicitation videos.

3.3. Biosignal sensors

Electrocardiographic (ECG) data acquisition was performed with a custom, two lead differential sensor design with virtual ground, proposed in [19]; Table 3 outlines the main specifications. Two ECG sensors were used in our experimental setup, one for signal acquisition at the hand palms with dry

Ag/AgCl electrodes, and another for signal acquisition with Electrolycras at the index and middle fingers. For improved comfort and greater efficiency, the ECG sensors were fitted to a leather base, with the intended hand placement signaled in an unequivocal way.

Fig. 3 depicts the devised sensor and electrodes arrangement. One of the ECG sensors was connected to the dry Ag/AgCl electrodes that were placed at the base of the hand palms, near the thenar eminence; the other ECG sensor was connected to the Electrolycra strips placed along the index and middle fingers. The transfer function for this sensor is given by Eq. (1); V_O is the output voltage of the sensor and V_{ECG} is the ECG signal voltage, both in *Volt*. Eq. (2) provides an equivalent representation, which uses the digital output codes D_O of the ADC directly. In our setup we used a system with 12-bit resolution ($n = 12$), $V_{cc} = 5$ V and $V_{ss} = 0$ V.

$$V_{ECG} = \frac{V_O}{1000} \tag{1}$$

$$V_{ECG} = \frac{D_O * V_{cc}}{2^n \times 10^3} \tag{2}$$

As previously mentioned, simultaneous acquisition of EDA data was also performed to assess the arousal level of each participant. We recurred to the commercially available edaPLUX,[6] which has two independent sensor leads, allowing a high level of flexibility in the sensor placement. Pre-gelled and self-adhesive Ag/AgCl electrodes[7] were used as interface with the skin for improved conductivity. Table 3 outlines the main specification of the sensor, and Fig. 2 depicts the sensor and electrodes.

The transfer function for this sensor is given by Eq. 3; V_O is the output voltage of the sensor in *Volt*, and G_S is the skin conductance in μS (micro Siemens). Eq. 4 provides an equivalent representation, which uses the digital output codes D_O of the ADC directly. In our setup we used a system with 12-bit resolution ($n = 12$), $V_{cc} = 5$ V and $V_{ss} = 0$ V.

$$G_S = \frac{V_O}{0.2} \tag{3}$$

$$G_S = \frac{D_O \times V_{cc}}{2^n \times 0.2} \tag{4}$$

3.4. Data acquisition

Raw biosignals were acquired with a bioPLUX research,[8] Bluetooth wireless biosignal acquisition unit; this device was

---

[2] http://www.youtube.com/watch?v=wwO-wo892pI.
[3] http://www.youtube.com/watch?v=zea6SCzjtfw.
[4] http://www.youtube.com/watch?v=0_bx8bnCoiU.
[5] http://www.youtube.com/watch?v=YGJ_jPKOj1c.

[6] http://www.biosignalsplux.com.
[7] http://www.spesmedica.it.
[8] http://www.biosignalsplux.com.

| Table 4 – Specifications of the bioPLUX research.	
Connectivity	Class II Bluetooth connectivity
Range	up to 10 m
Resolution	up to 12 bit
Sampling rate	up to 1000 Hz
Weight	74 g
Size	84 mm × 53 mm × 18 mm
Battery	Li-On; 7.4 V; 800 mAh

Fig. 4 – Examples of the signals extracted from our initial experiments. All signals are normalized for improved legibility. On the top chart we show the LED+TTL Switch synchronization signals, while the ECG signals collected with both electrode materials are shown in the bottom chart.

used in a 12-bit resolution, 1 kHz sampling frequency configuration. To guarantee electrical isolation between both ECG sensors used in the experiments, two independent biosignal acquisition units were used. Table 4 describes the main specification of the biosignal acquisition unit.

Fig. 8 depicts the hardware configuration used in our experiments for biosignal acquisition. As previously described, an iPad was used for to reproduce the multimedia contents presented during the computer task/acquisition; two Light Dependent Resistor (LDR) sensors were taped with black tape (for discreetness purposes), to the corner of the iPad screen above the area where the video-embedded white square would appear. The ECG sensor arrangement was placed in front of the iPad, in such way that the person could comfortably relax the hands while performing the computer task, that is, watch the video clips. Electrodermal activity sensors were placed near the ECG sensor arrangement to enable fast application to the person. All the sensors were connected to each of the biosignal acquisition units that were placed close to the remaining apparatus. Table 5 presents the channel mapping for each unit.

3.5. Time synchronization

Time synchronization between both biosignal acquisition units was performed optically to ensure full electrical decoupling, using elements from a syncPLUX synchronization kit. To one of the units, we connected a switch, which simultaneously activated a light emitting diode (LED), and triggered a TTL signal to the digital input port of the device; to the other unit, we connected a LDR, which was directly attached to the LED. With this setup, whenever the switch was pressed, a common signal was recorded by both devices, enabling time alignment in the post-processing (Fig. 4).

A second synchronization mechanism was also implemented using LDR sensors, to guarantee synchronization between the biosignal acquisition units and the multimedia content displayed. For this purpose, two LDR sensors were applied to the iPad screen and connected to each of the biosignal acquisition units. In a pre-test stage, the videos were played to capture a "videoprint" based on the screen light intensity variations, which can be used to relate the biosignals activity with the stimuli being presented.

Furthermore, this mechanism allowed us to introduce redundancy in the synchronization process, as the information extracted from the videoprints alone can be used for time alignment of the signals collected by each independent unit. Since the operating principle of the LDR sensors is also optical, electrical decoupling between both independent biosignal acquisition units is once again guaranteed.
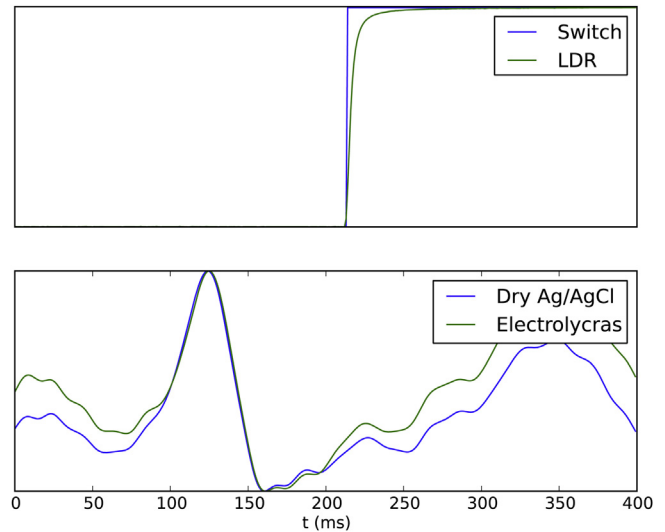
4. Short-term dataset

4.1. Participants

The experimental setup was placed on an unrestrained setting, in the lobby of the building where one of our research teams is located: Torre Norte of the Alameda Campus of Instituto Superior Técnico. To generate awareness for the initiative and increase compliance, a set of promotional banners was created and placed close to the workbench; a volunteer counter, was regularly updated to further motivate potential candidates. Fig. 11 shows the devised designs, seen also in their intended use in Fig. 1(a).

Sessions took place over two full days, and in the end data had been collected for an overall total of 65 participants, the majority of which were engineering students and researchers. The demographics shown 49 males and 16 females, with an average age of 31.1 ± 9.46 years. None of the participants reported any health problems, reason for which we consider the collected data to be representative of the healthy population.

As a reward for participating in the experiments, a gift card created from the collected biosignals was prepared (Fig. 10). With the individual heartbeat waveforms, an orange colormap image was created, where each line corresponds to an individual heartbeat waveform, and each column corresponds to the amplitude of each sample of the heartbeat wave data. With the individual EDA events, a blue colormap image was created using the same principle, but in this case rotated of −90°.

Table 5 – Channel mappings and sensor references.		
Ch.	Sensor	Placement
00:07:80:42:0F:8B bioPLUX research		
1	ECG-3-82	Hand palms (with Ag/Cl electrode and no gel)
2	EDA-1-25	Left hand
3	LDR-1-4	iPad screen
4	LDR-1-7	LED
00:07:80:42:0F:85 bioPLUX research		
1	ECG-3-83	Index and middle fingers (with electrolycra)
5	LDR-1-1	iPad screen
i/o	SYNC	Sync adapter + LED (in direct contact with LDR-1-7)

4.2.   *Experimental procedure*

The whole experiment was designed to have an average total completion time of approximately 5 min, scattered throughout three stages: (a) Informed consent (CI); (b) Low arousal video (A1); and (c) High arousal video (A2). Once participants willingly showed interest in becoming part of the experiments, the first stage would take place, and consisted on having the coordinator of the experiment going through the informed consent (Fig. 9), explaining the procedure in detail, goals of the study, and related background work. During this phase, only ECG signals at the hand palms and fingers were recorded, and no emotion eliciting stimuli were presented.

For the second phase, subjects were fitted with the EDA sensor on the left hand, in which one terminal of the sensor was placed on the middle phalanx of the index finger, and the other on the middle phalanx of the ring finger. The headphones were placed over the ears, and the low arousal video was started on the iPad by the coordinator of the experiment. For the third phase, the headphones were kept on the subject, and the video on the iPad switched to the high arousal video sequence. During both these phases, ECG signals at the palms, and EDA signals were recorded.

At each stage, recordings from each individual device were stored in a separate file for more efficient post-processing, and labeled with the date, identification of the system, and a code assigned to the subject. In the beginning and in the end of each recording the switch would be triggered in order to produce the synchronization signal, necessary to find the common time base as previously described (Fig. 4).

5.   Long-term dataset

5.1.   *Participants*

The long-term dataset was collected over a period of several days and different settings, reason for which a slimmed-down version of the proposed acquisition setup was adopted. We used a single biosignal acquisition system with just one sensor to measure ECG fitted with the dry Ag/AgCl electrodes. The experimental setup was prepared at the cardiopneumology laboratory located at the Escola Superior de Saúde da Cruz Vermelha Portuguesa, that was prepared to receive the participants.

The recruitment of volunteers was performed by advertising the days in which data acquisition sessions were held, and by providing an overview of the action during classes. A total of 63 subjects (nursing and health technologies students), were enrolled in the experiment and participated in the two acquisition sessions. The demographics showed 14 males and 49 females, with an average age of 20.68 ± 2.83 years. None of the participants reported any health problems, reason for which we consider the collected data to be representative of the normal population.

5.2.   *Experimental procedure*

Two data acquisition moments separated by a 3-month time-frame took place, and enabled data recollection from the set of volunteers enrolled in the experiment, with the purpose of studying the changes in the ECG morphology over time. In the first moment, the coordinator of the experiment presented the informed consent to prospective participants (Fig. 9), explaining the procedure in detail, goals of the study, and related background work. Once participants willingly showed interest in becoming part of the experiments, they were required to sign the document and were enrolled in the data collection program.

In both moments only ECG signals at the fingers were recorded, and in each of the sessions the subjects were asked to sit for 2 min in a resting position, with two fingers, one from the left and another from the right hand, placed in each of the dry Ag/AgCl electrodes. The recordings were stored in individual files for more efficient post-processing, and labeled with the date, identification of the system, and a code assigned to the subject. The fact that it has two sessions separated by several months apart makes this dataset particularly useful for benchmarking the performance of identity recognition algorithms taking into account potential variations in the heartbeat waveform over time.

6.   Status report

6.1.   *Outline*

Our experiments to date have already enabled several interesting findings using the collected data. First, as shown in Fig. 4 (bottom), the optical synchronization method enables time alignment within the millisecond, allowing the signals collected by each independent unit to be compared in the same time base. Also, as we can see from the example presented in Fig. 5, both the ECG signals collected with dry Ag/AgCl and Electrolycras at the hand palms and at the fingers retain a high level of detail with respect to the heartbeat waveform.
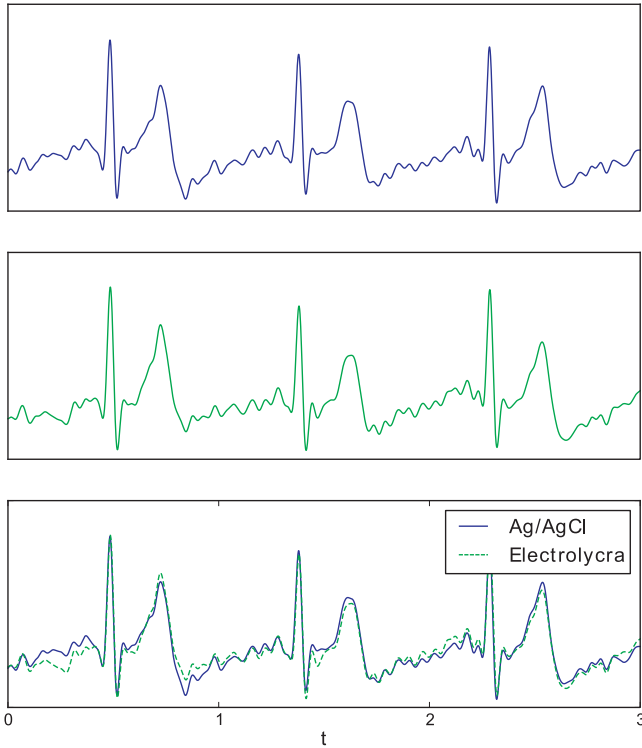
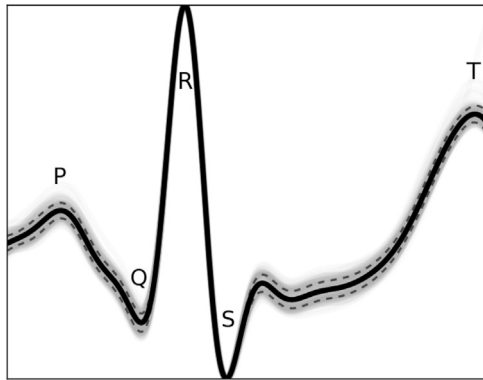Fig. 5 – ECG signals at the hand palms (dry Ag/AgCl) and fingers (Electrolycras).

Fig. 6 – Example of a filtered and segmented ECG signal collected at the hand palms with dry Ag/AgCl. The black line depicts the average of all waveforms, while the dashed lines depict the standard deviation; the gray lines show an overlay of the segmented individual heartbeat waveforms for one of the acquisition sessions.

In particular, not only the so-important QRS complex is clearly noticeable, but also the P and T waves can be found. Moreover, we can see that there is a high morphological similarity between both signals, leading us to believe that both the electrolycras and the dry Ag/AgCl electrodes can be used as interface with the skin, giving higher flexibility and convenience to the biometric system designer. This is further reinforced by the analysis performed in [8,14].

(a) A1 video.   (b) A2 video.

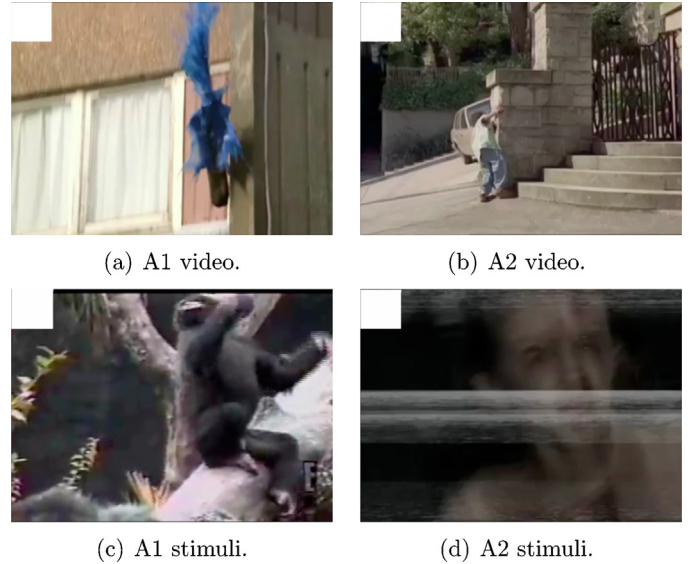(c) A1 stimuli.   (d) A2 stimuli.

Fig. 7 – Sample frames from the video and stimuli sequences.

Fig. 8 – Full experimental setup.

An additional example of the collected signals is presented in Fig. 6, where we can see an example of a filtered and segmented recording collected at the hand palms with the dry Ag/AgCl electrodes.

Still, signals collected at the hands and fingers, especially using dry materials as electrodes, have a lower signal-to-noise ratio, which raises several challenges namely in terms of filtering, segmentation, and outlier detection. A first set of guidelines regarding the pre-processing of these signals can be found in the work by Lourenço et al. [9,10]; more recently, to enhance the ECG signal quality and increase the signal-to-noise ratio in the ECG biometrics context, we started to adopt a 300 order band pass Finite Impulse Response (FIR) filter, with a Hamming window, and cutoff frequencies of 5 Hz to 20 Hz. These specifications take into account the ECG information bandwidth and empirical considerations obtained from the data itself.

For researchers working on fiducial or partially fiducial approaches, segmentation is a key aspect in terms of data analysis. We refer the reader to the papers by Bachler et al. [20]

Informed Consent Form

We are performing a study on biometric recognition and emotional responses based on biosignals.

Your participation in our study is fundamental to help expand the knowledge in the field.

★★★★★
http://www.it.pt

Sensors   Data   Results   +You   more ▶

Materials

An electrocradiography (ECG) sensor will be used for the biometric recognition part.

An electrodermal activity (EDA) sensor will be used for the emotional response part.

A computer will be used to elicit emotions.

Sensors   more ▼

Objectives

Assess the quality of ECG signals at the fingers.

Evaluate electro-textiles for ECG acquisition.

Characterize the skin conductance levels.

Preliminary analysis of EDA laterality.

Results   more ▼

Methods

All sensors are placed at the finger or hand level.

All sensors only have a measurement function.

All acquired signals will be used anonimously.

You can decide to interrupt the study at anytime.

Data   more ◀

About You

Name:

E-Mail: Age:

+You   more ▶

Thank's!

We greatly appreciate your time and support.

If you have additional questions about the study contact us:
afred@lx.it.pt | arlourenco@lx.it.pt | hsilva@lx.it.pt

more ▼

Signature

I acknowledge to have been informed about this study and accept to participate as a volunteer:

_____/____/____

Notes and Comments

Fig. 9 – Informed consent document that was read to and signed by the volunteers.

and by Canento et al. [21], in which a description and comparison of several ECG segmentation algorithms is presented.

6.2.   Studies with CYBHi

In part or in full, this dataset has already been used in several peer reviewed and accepted state-of-the-art publications. In [13] a study of different segmentation algorithms is performed, which addresses the problem of online vs. offline segmentation of hand/fingers ECG data; 63 records from the CYBHi short-term dataset were used, and the authors have concluded that online approaches present competitive results both in terms of the number of overall detected segments, and on the percentage of those that can be considered to be valid.

In [12], a study on the biometric performance of full vs. partial waveform is presented. A total of 32 records from the CYHBi short-term dataset were used, to evaluate how the use of the waveform comprehending the P-QRS-T segments compares to using only the R-ST portion of the waveform. Experimental evaluation was performed both on data from the
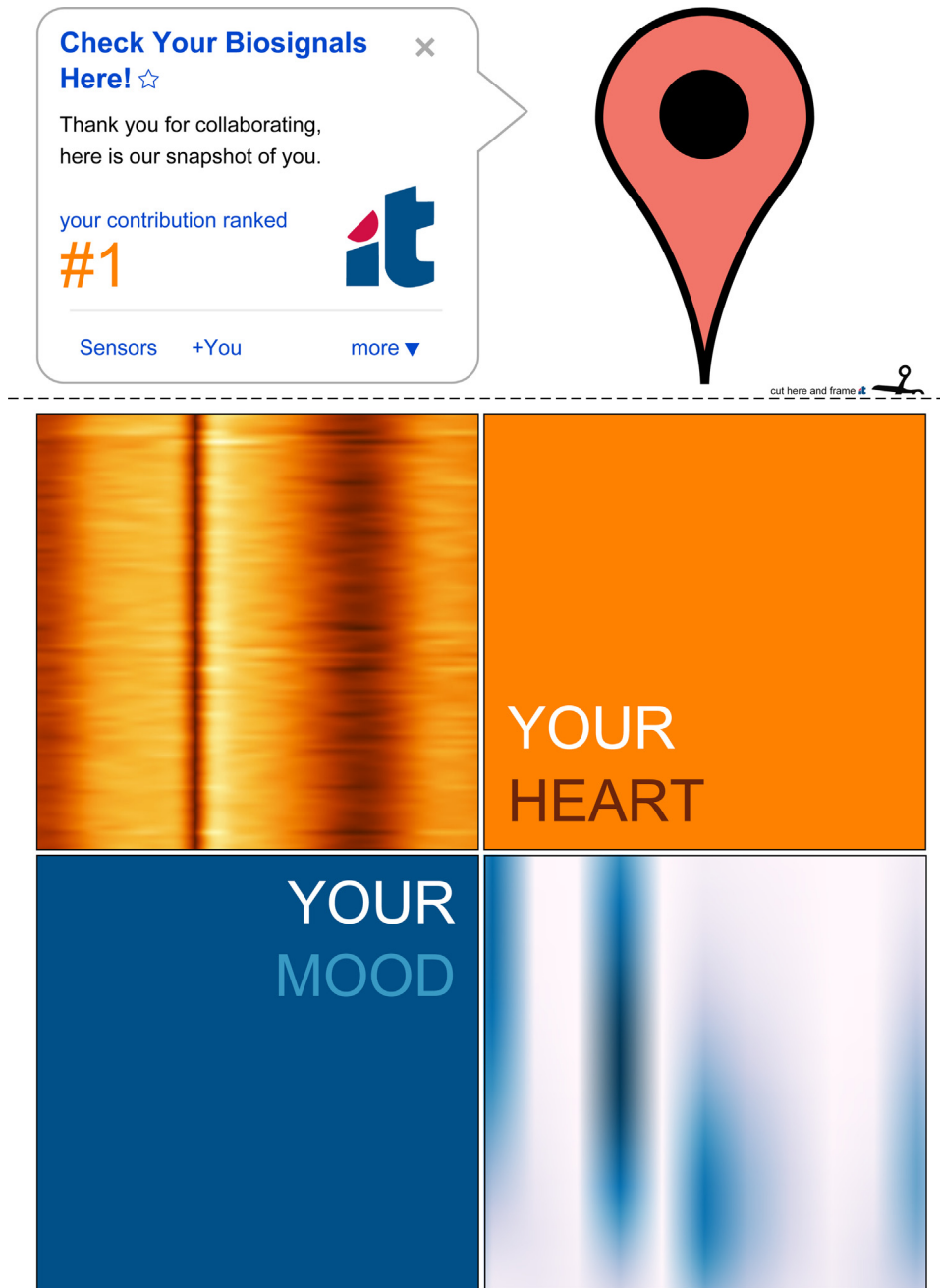
Fig. 10 – Participant gift card offered to volunteers enrolled in the experiments, showing highlights of their own signals.

dry Ag/AgCl electrodes and Electrolycras, and results revealed that the P-QRS-T provides higher recognition rates due to the additional information; in the best case scenario a $1.66\% \pm 2.28$ identification error (EID) was achieved for groups of 2 users. Nonetheless, with a marginal decrease in performance, the RS-T segment based approach is able to provide an adequate compromise solution, given that it has a lower temporal and spacial complexity; for groups of 2 users, this approach has revealed $2.40\% \pm 3.96$ EID.

Targeting the real-time classification scenario, in [9] a total of 32 records from the short-term CYBHi were used, to compare the performance of a Nearest Neighbor classifier with an approach based on Support Vector Machines (SVM). Experimental results have shown that, depending on the application, each approach has its own advantages, and that the SVM based approach enables high recognition rates, for example, in an authentication task, for an operating point where the False Acceptance Rate (FAR) is 0%, the False Rejection Rate (FRR) can be as low as $13.91\% \pm 4.55$.

In [22], two PCA-based methods are presented, which use the full CYBHi short-term dataset to explore the concepts of Individual Eigen-Heartbeat and Overall Population Eigen-Heartbeat, as a method of feature extraction and template creation. Results have shown that a 0% EID can be achieved.
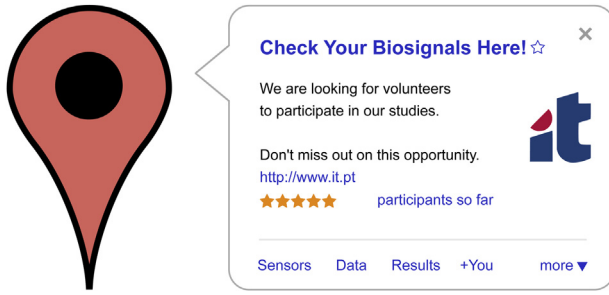
Fig. 11 – Promotional banners created to signal the initiative to prospective participants.

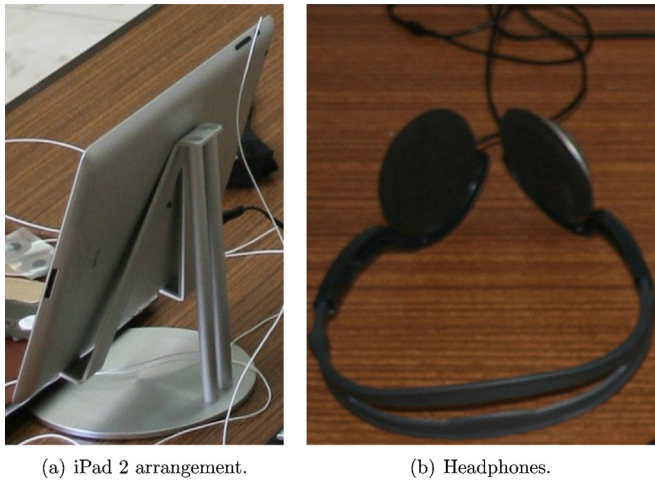(a) iPad 2 arrangement.   (b) Headphones.

Fig. 12 – Multimedia setup for emotion elicitation.

In [23,24], the authors tried to cope with outliers and time-variance introduced by acquisitions over time using wavelets and clustering-based methods. Using 63 records, and the two temporal separated sessions from the long-term CYBHi dataset, Equal Error Rates (EER) of 11.5% and 12.4% were achieved, revealing that this is a challenging task. More recently in [11], we further evaluated the temporal stability of ECG signals collected at the fingers, reaching an EER of 9.10% using SVMs also for the CYBHi long-term dataset.

7. Resources

7.1. Materials

All materials and outputs of the data acquisition sessions were grouped into a folder structure, which can be used for future

Table 7 – Acquired data file structure.

Col.	Content
1	7-bit sample sequence number (from the device)
2	Digital input (from the device)
3	Digital output toggle (from the software)
≥4	Analog channels ≥1 (from the device)

reference. These include graphic materials, videos, the collected data, and support documentation; Table 6 outlines what can be found in each folder.

7.2. Data format

Each file contains the data collected from the biosignal acquisition and synchronization sensors. Included data comprises the modalities described in Section 3; the different moments of the experimental procedure acquired with the two biosignal acquisition units are stored in separate files. Files are identified with indication of the date, a code associated with the subject, and the moment to which they correspond (e.g. YYYYMMDD-AAA-BB-CC.txt), the general notation being given by:

```
<filename>      ::=<date> "–" <code> "–"
                   <moment> "–" <unit> ".txt"
<date>          ::= YYYYMMDD
<code>          ::=up to three characters
<moment>        ::="CI"|"A0"|"A1"|"A2"
<unit>          ::="8B"|"85"
```

In the above notation used to distinguish the moment, "CI" denotes the informed consent part of the test, while "A1" and "A2" denote respectively the low and high arousal videos in the short-term dataset; "A0" denotes the acquisitions of the long-term dataset. In the notation used to distinguish the biosignal acquisition unit, "8B" and "85" refer to the "00:07:80:42:0F:8B" and "00:07:80:42:0F:85" units, respectively. Data is stored in standard ASCII text with a bioPLUX file format, where each line corresponds to one sample; the columns mapping is described in Table 7. The header has eight lines, preceded by the # symbol, and the stored information is the following:

```
# bioPlux Text File Format
# Version: 1
# StartDateTime: YYYY-MM-DD HH:MM:SS
# SamplingFrequency: 1000
# SampledChannels: 1 2 3 4 5 6 7 8
# SamplingResolution: 12
# AcquiringDevice: XX:XX:XX:XX:XX:XX
# EndOfHeader
```

Table 6 – Description of the project resources.

Folder	Content
data	*Raw* data collected during the experiments stored in individual *.txt* files
documents	Support documentation used in the experiments: informed consent, participant list, and acquisition setup specification
graphics	Graphic materials prepared for the experiments: banners and gift cards
photos	Pictures of different aspects of the experimental setup
results	Graphics and data files with results from initial data processing
stimuli	Film clips presented as the computer task to the subjects

8. Lessons learned

While datasets with ECG collected at the chest using clinical-grade equipment are readily available, the same cannot be said regarding signals collected at the hand palms or fingers in an off-the person approach. Among other potential applications, these unconventional locations are fundamental for the growing research field of ECG biometrics, as the chest placement is far too intrusive for feasible real-world deployment. These novel trends in ECG biometrics require more convenient signal acquisition methods however, in this particular area, researchers still lack publicly available datasets that can promote benchmarking and comparison of results across teams.

The Check Your Biosignals Here initiative (CYBHi) dataset, provides high resolution raw data from signals collected at the hand palms and fingers in an off-the-person approach, using dry Ag/AgCl electrodes and Electrolycras as interface with the skin. We have described a biosignal acquisition framework, devised to overcome the need for a repeatable and easy-to-apply procedure for data collection in large population groups. This setup builds upon prior work from our group in the area of biosignal research, as a tool to identify potentially time variant properties in the signals, and to evaluate the robustness of algorithms and methods to changes in the environmental conditions and other data acquisition influencing factors.

We started to compile a short-term dataset where a single acquisition session was performed to collect signals that can compare different electrode materials, as well as the effect of arousal-eliciting stimuli in the heartbeat waveform templates. A long-term dataset was also started, in which two data acquisition sessions were performed, enabling the evaluation and benchmarking of the persistence of the heartbeat waveform templates over time. Several studies have already used the CYBHi dataset and provided valuable insight on the potential of the collected data, as well as guidelines on how to perform pre-processing of the data. Our dataset is already made available for researchers worldwide.[9] Furthermore, we also created a set of tools that researchers can use to further contribute to the CYBHi dataset, and replicate the experiments, in particular, by providing access to the biosignal acquisition hardware itself.[10]

9. Future plans

Ongoing work is focusing on further extending the number of subjects both in the short-term and long-term datasets, and also enriching the potential of our dataset by recalling the subjects that were previously assessed in order to have a record of the cardiac activity of each subject in different moments in time separated by several months or years apart. We are currently recalling the population enrolled in the CYBHi long-term experiments, in order to record 1-year apart data.

In parallel, we are also working to provide morphological ground-truth data by manual labeling of fiducial points in the recordings, which will extend the use of our dataset

beyond the field of ECG biometrics, enabling for example the assessment and benchmarking of filtering and segmentation algorithms [25–27]. Another aspect that will be analyzed in future research work is the correlation between the ECG signals and the variations of the arousal level of each subject as measured through the collected EDA data.

Regarding the dataset itself, we will seek the possibility of making it available through Physionet [7] in a near time, and making it compliant with the ISO standards for biometric data storage and exchange. Finally, the dataset is now provided as standard ASCII text files; future work will also focus on migrating the dataset to a structured database format [28–30].

REFERENCES

[1] A.K. Jain, A. Ross, K. Nandakumar, Introduction to Biometrics, Springer US, New York, USA, 2011.

[2] Y. Wang, F. Agrafioti, D. Hatzinakos, K.N. Plataniotis, Analysis of human electrocardiogram for biometric recognition, EURASIP Journal on Advances in Signal Processing 2008 (2008) 19.

[3] A. Chan, M.M. Hamdy, A. Badre, V. Badee, Wavelet distance measure for person identification using electrocardiograms, IEEE Transactions on Instrumentation and Measurement 57 (2) (2008) 248–253.

[4] I. Odinaka, P.-H. Lai, A.D. Kaplan, J.A. O'Sullivan, E.J. Sirevaag, J.W. Rohrbaugh, ECG biometric recognition: a comparative analysis, IEEE Transactions on Information Forensics and Security 7 (6) (2012) 1812–1824.

[5] S.A. Israel, J.M. Irvine, B.K. Wiederhold, M.D. Wiederhold, The Heartbeat: The Living Biometric, Wiley-IEEE Press, Hoboken, New Jersey, 2009, pp. 429–459 (Chapter 17).

[6] C. Zhao, T. Wysocki, F. Agrafioti, D. Hatzinakos, Securing handheld devices and fingerprint readers with ECG biometrics, in: Proc. of the IEEE Int'l Conf. on Biometrics: Theory, Applications and Systems (BTAS), 2012, pp. 150–155.

[7] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, H. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) 215–220.

[8] H. Silva, C. Carreiras, A. Lourenço, A. Fred, Off-the-person electrocardiography, in: Proc. of the Int'l Congress on Cardiovascular Technologies (CARDIOTECHNIX), 2013, pp. 99–106.

[9] A. Lourenço, H. Silva, A. Fred, ECG-based biometrics: a real time classification approach, in: Proc. of the IEEE Int'l

---

9 http://camoes.lx.it.pt/CYBHi.rar.
10 http://www.bitalino.com.

Workshop on Machine Learning for Signal Processing (MLSP), 2012, pp. 1–6.

[10] A. Lourenço, H. Silva, A. Fred, Unveiling the biometric potential of Finger-Based ECG signals, Computational Intelligence and Neuroscience 2011 (2011) 5.

[11] H. Silva, A. Lourenço, A. Fred, A. Jain, Finger ECG signal for user authentication: usability and performance, in: Proc. of the IEEE Int'l Conf. on Biometrics: Theory, Applications and Systems (BTAS), 2013.

[12] H. Silva, A. Lourenço, A. Fred, In-vehicle driver recognition based on hand ECG signals, in: Proc. of the ACM Int'l Conf. on Intelligent User Interfaces (IUI), 2012, pp. 25–28.

[13] A. Lourenço, H. Silva, P. Leite, R. Lourenço, A. Fred, Real time electrocardiogram segmentation for finger based ECG biometric, in: Proc. of the Int'l Conf. on Bio-Inspired Systems and Signal Processing (BIOSIGNALS), 2012, pp. 49–54.

[14] C. Carreiras, A. Lourenço, H. Silva, A. Fred, Comparative study of medical-grade and off-the-person ECG systems, in: Proc. of the Int'l Congress on Cardiovascular Technologies (CARDIOTECHNIX), 2013, pp. 115–120.

[15] C. Stickel, M. Ebner, S. Steinbach-Nordmann, G. Searle, A. Holzinger, Emotion detection: application of the valence arousal space for rapid biological usability testing to enhance universal access, in: C. Stephanidis (Ed.), Universal Access in Human–Computer Interaction. Addressing Diversity, Vol. 5614 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2009, pp. 615–624.

[16] R.W. Picard, Affective Computing, The MIT Press, Cambridge, USA, 2000.

[17] F. Agrafioti, D. Hatzinakos, A.K. Anderson, ECG pattern analysis for emotion detection, IEEE Transactions on Affective Computing 3 (1) (2012) 102–115.

[18] C. Carreiras, H. Aidos, H. Silva, A. Fred, Exploratory EEG analysis using clustering and Phase-Locking factor, in: Proc. of the Int'l Conf. on Bio-inspired Systems and Signal Processing (BIOSIGNALS), 2013, pp. 79–88.

[19] H. Silva, A. Lourenço, R. Lourenço, P. Leite, D. Coutinho, A. Fred, Study and evaluation of a single differential sensor design based on electro-textile electrodes for ECG biometrics applications, in: Proc. of the IEEE Sensors Conf., 2011, pp. 1764–1767.

[20] M. Bachler, C. Mayer, B. Hametner, S. Wassertheurer, A. Holzinger, Online and offline determination of QT and PR interval and QRS duration in electrocardiography, in: Q. Zu, B. Hu, A. Eli (Eds.), Pervasive Computing and the Networked World, Vol. 7719 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2013, pp. 1–15.

[21] F. Canento, A. Lourenço, H. Silva, A. Fred, On real time ECG segmentation algorithms for biometric applications, in: Proc. of the Int'l Conf. on Bio-inspired Systems and Signal Processing (BIOSIGNALS), 2013, pp. 228–235.

[22] M. Santos, A. Fred, H. Silva, A. Lourenço, Eigen heartbeats for user identification, in: Proc. of the Int'l Conf. on Bio-inspired Systems and Signal Processing (BIOSIGNALS), 2013, pp. 351–355.

[23] C. Carreiras, A. Lourenço, H. Silva, A.L.N. Fred, A unifying approach to ECG biometric recognition using the wavelet transform, in: Proc. of the Int'l Conf. on Image Analysis and Recognition (ICIAR), 2013, pp. 53–62.

[24] A. Lourenço, H. Silva, C. Carreiras, A.L.N. Fred, Outlier detection in non-intrusive ECG biometric system, in: Proc. of the Int'l Conf. on Image Analysis and Recognition (ICIAR), 2013, pp. 43–52.

[25] J. Abawajy, A. Kelarev, M. Chowdhury, Multistage approach for clustering and classification of ECG data, Computer Methods and Programs in Biomedicine 112 (3) (2013) 720–730.

[26] S. Das, K. Maharatna, Fractional dynamical model for the generation of ECG like signals from filtered coupled Van-der Pol oscillators, Computer Methods and Programs in Biomedicine 112 (3) (2013) 490–507.

[27] T. Ji, Q. Wu, Broadband noise suppression and feature identification of ECG waveforms using mathematical morphology and embedding theorem, Computer Methods and Programs in Biomedicine 112 (3) (2013) 466–480.

[28] C. Carreiras, H. Silva, A. Lourenço, A. Fred, StorageBIT: a metadata-aware, extensible, semantic, and hierarchical database for biosignals, in: Proc. of the Int'l Conf. on Health Informatics (HEALTHINF), 2013, pp. 65–74.

[29] K.K.-Y. Lee, W.-C. Tang, K.-S. Choi, Alternatives to relational database: comparison of NoSQL and XML approaches for clinical data storage, Computer Methods and Programs in Biomedicine 110 (1) (2013) 99–109.

[30] J. Meyer, S. Ostrzinski, D. Fredrich, C. Havemann, J. Krafczyk, W. Hoffmann, Efficient data management in a large-scale epidemiology research project, Computer Methods and Programs in Biomedicine 107 (3) (2012) 425–435.