Towards the Development of Sensor Platform for Processing Physiological Data From Wearable Sensors

Krzysztof Kutt, Wojciech Binek, Piotr Misiak, Grzegorz J. Nalepa, and Szymon Bobek

AGH University of Science and Technology
al. Mickiewicza 30, 30-059 Krakow, Poland
{kkutt,wojciech.binek,pmisiak,gjn}@agh.edu.pl

Abstract. The paper outlines a mobile sensor platform aimed at processing physiological data from wearable sensors. We discuss the requirements related to the use of low-cost portable devices in this scenario. Experimental analysis of four such devices, namely Microsoft Band 2, Empatica E4, eHealth Sensor Platform and BITalino (r)evolution is provided. Critical comparison of quality of HR and GSR signals leads to the conclusion that future works should focus on the BITalino, possibly combined with the MS Band 2 in some cases. This work is a foundation for possible applications in affective computing and telemedicine.

1 Introduction

Recent rapid development of wearable devices equipped with physiological sensors is an opportunity for new systems in the areas of telemedicine, quantified self, and affective computing (AfC). Our recent works [?] focus on the last of these areas. There are two main aspects of AfC [?]. The first one is related to the detection of emotional responses of humans. The second one is related to the simulation of emotional responses in artificial systems. We are interested in the first aspect, which recently can largely benefit from wearable devices.

In AfC appropriate identification of the affective condition of a person requires a certain model of emotions. There are multiple models considered in psychology, philosophy and cognitive science. William James was the precursor of the appraisal theory which is among most popular in the community of computational emotional modeling [?,?]. One of the most popular appraisal theories is OCC (Ortony, Clore & Collins) [?] which categorizes emotion on basis of appraisal of pleasure/displeasure (valence) and intensity (arousal). Research indicates that they can be measured by the use of Autonomic Nervous System (ANS) activity, including the use of Heart Rate (HR) and Skin Conductance/Galvanic Skin Response (GSR) signals (for meta-analysis see [?]).

The ANS measures can be accurately acquired in laboratory experiments. However, a practical challenge is the quality of measurement provided by field devices, such as wearables, e.g. wristbands. These devices are low-cost and accessible thus creating opportunity for real life applications. On the other hand they use lower quality sensors, often applied in a non-optimal way. Recently there has been growing interest in assessing the quality of such devices, e.g. [**?**].

The original contribution of our work presented in this paper is the critical evaluation of several wearable devices delivering physiological data monitoring. We aim at comparing these devices considering future applications in AfC and in telemedicine. As such, we focus on the continous monitoring of the HR and GSR signals. The rest of the paper is organized as follows. We begin with the overall design of the sensor platform in Sect. 2, then moving to the discussion of selected devices in Sect. 3. Based on this, we present the measurement procedures in Sect. 4, along with the detailed signal processing in Sect. 5. We then move to the evaluation of results in Sect. 6 and conclusions in the final section.

2 Outline of a Mobile Sensor Platform

Our proposal of the mobile platform aimed at Affective Computing applications supports the affective data flow through several interconnected modules:

1. The person is experiencing emotion which is connected with the reactions of person's ANS.
2. Mobile sensor monitors these signals (e.g. HR and GSR).
3. Data is transmitted to the processing device (e.g. smartphone) via bluetooth or other interface.
4. The processing device reads the data using API provided by the mobile sensor distributor.
5. Statistical or machine learning model is used to transform sensor data into emotion values (e.g. in Valence x Arousal dimensions or nominal values defining the names of emotions).
6. All data, i.e. gathered from sensors and outputs from model, may be saved in CSV files, broadcasted to other applications or combined with other data streams (e.g. GPS signal and network connection usage) to provide more reliable contextual information[1].
7. Finally, the data may be used by number of applications, including: affect identification, context processing and adaptation, and health monitoring.

Our work presented in this paper is focused on steps 2-4, i.e. on gathering sensory data from user. With this in mind, one can specify requirements for mobile sensors to be fit in the presented platform:

1. The ANS measurement should be *accurate*. As we aim at low-cost devices available for almost everyone, it cannot be very precise. There is only a need for differentation of various valence and arousal levels and their changes, what makes the devices usable from the AfC applications point of view.

[1] With the use of e.g. AWARE framework, see `http://www.awareframework.com/`.

2. Collecting affective information should be done on a continuous basis, which is related to: (a) platform *mobility*, as it will assist user everywhere, (b) reliable sensor *contact*, as it will assist user during various activites, (c) being *comfortable* for user, as it should not distract her in regular life, (d) sufficient *battery* capacity that lasts at least one working day without recharging.
3. Data should be processed live to provide an affective feedback loop, so there is a need for: (a) *connection* with mobile device, e.g. through Bluetooth, (b) *raw signal* access, as each filtering done by the sensor or API results in data loss, what makes further processing difficult, (c) clearly defined *unit of measurement*, to provide a possibility of comparison with other devices and to allow the use of general model that gathers signals in some specified units, (d) *open API* that will allow the access to current sensor readings.

Low-cost mobile sensors that we consider are described in the next section.

3 Overview of Selected Devices

Wristbands Empatica E4 [**?**] is a research- and clinical-oriented sensory wristband based on the technologies previously developed in the Affective Computing division of MIT Media Lab. The band has a photoplethysmography sensor for blood volume pulse measurements, as well as galvanic skin response sensor, infrared thermopile, 3-axis accelerometer and event mark button.

Microsoft Band 2 is a health and fitness tracking-oriented wristband. Equipped with optical heart rate and galvanic skin response sensors, as well as skin temperature, ambient light and UV sensors, 3-axis accelerometer, GPS and barometer.

Signals from both Empatica E4 and MS Band 2 were obtained and recorded via a custom dedicated application for Android devices created by authors [**?**].

BITalino The BITalino (r)evolution kit[2] is a complete platform designed to deal with the body signals. It is ready to use out-of-the-box and allows the user to acquire biometric data using included software, which also enables the full control of the device. The device sends raw signals produced by the analog-to-digital converter. These can be converted to the correct physical units using the right transfer function for the sensor that produced the data. Communication is available via Bluetooth or any UART-compatible device (e.g. ZigBee, WiFly or FTDI). There is also set of APIs for many platforms including Arduino, Android, Python, Java, iOS and many more, which let the user create custom software.

e-Health The e-Health Sensor Platform V2.0[3] is an open-source platform that allows users to develop biometric and medical applications where body monitoring is used. It is possible to perform real time monitoring or to get sensitive data, which will be subsequently analysed for medical diagnosis. Our e-Health kit consists of the following components useful for AfC experiments:

[2] For details see `http://bitalino.com/`

[3] For details see `https://www.cooking-hacks.com/documentation/tutorials/ehealth-biometric-sensor-platform-arduino-raspberry-pi-medical`

- e-Health PCB, which can be connected to Arduino or Raspberry Pi; all sensors are connected to this PCB,
- heart rate/blood saturation sensor – easy to use on-finger sensor,
- electrocardiogram (ECG) – monitors the electrical and muscular functions of the heart. It consists of 3 electrodes and 3 leads (positive, negative, neutral),
- body temperature sensor,
- galvanic skin response (GSR/EDA) – the sensor consists of two metallic electrodes and is a type of ohmmeter with human body being a resistor,
- electromyograph (EMG) – measures the electrical activity of skeletal muscles by detecting the electrical potential generated by muscle cells when these cells are activated. It consists of 3 electrodes (MID, END and GND).

The e-Health library methods are needed to get measured data. In most cases the value is passed using Arduino analog input pins connected to the responding output pins of the e-Health PCB. These methods read the voltage and use it to calculate the value returned by the method. Serial communication is the best way to transfer data, but wireless channels like WiFi, Bluetooth are also available.

4 Evaluation and Measurement Procedures

To achieve the goal of devices comparison with consideration of future AfC applications, experimental procedure consisting of 3 parts was designed. It was aimed at data acquisition in various settings similar to target Affective Computing setting, including various affective states that cover wide range of valence and arousal values. Affect-related physiological signals, Heart Rate (HR) and Galvanic Skin Response (GSR), were collected using four low-cost wearable devices described in Section 3 and Polar H6 strap as a reference. It is a professional fitness device used for HR tracking.

The first part was designed using the PsychoPy 2[4] environment, a standard software framework in Python to support psychological experiments. Subjects were asked to watch affective pictures. Each of them was presented for 3 seconds, then it disappears and subject had 5 seconds for valence evaluation on 7-levels scale $[1, 7]$. The set of 60 pictures was grouped into training session (6 images) and three experimental sessions (each of them with 18 images). To provide valid emotional descriptions of images, in terms of valence and arousal scores, a subset of Nencki Affective Picture System was used [?].

In the second part, the "London Bridge" platform game was run. The task was to collect points as you go through the 2D world. The gameplay incorporates current score and remaining time indicators, and random events of current score reduction or remaining time shortening. Full game design is discussed in [?].

Finally, the accuracy of acquired data during the physical activity was examined. After finishing the game, participants were asked to do 20 squats and rest for a minute. This should induce significant changes in both HR and GSR.

[4] See: http://psychopy.org.

5 Analysis Workflow

Performed experiment was focused on two basic parameters: HR and GSR. To evaluate HR signals we used Polar H6 chest strap as a reference. For GSR no reference data was available therefore different processing strategy was applied.

HR processing MS Band 2, eHealth and reference Polar H6 provide direct HR measurements. Polar H6 and MS Band 2 record HR with 1 Hz sampling rate, eHealth uses 32 Hz sampling rate. Because upsampling would not introduce any new information, we decided to downsample all signals to 1 Hz. E4 and BITalino do not provide direct HR information. The devices record blood volume pressure (BVP), and electrocardiography (ECG) signals that we used to calculate HR.

Proposed algorithm for HR detection is based on primary tone extraction methods. It combines autocorrelation and frequency domain analysis. Input signal is filtered using FIR bandpass filter preserving frequency range from 1.1 to 5 Hz. Filtered data is windowed using 8 s triangular, asymmetric window with 0.875 overlap producing an output of one sample per second. For each window position we calculate an autocorrelation and power spectrum. The first HR candidate is calculated from the delay between first and second maximum of autocorrelation result. Furthermore three candidates are calculated from position of first peak in power spectrum and distances between first, second and third peak. Both operations give total of 4 HR candidates. If previous HR sample is available, candidates that differ more than a defined threshold (7 BPM) are discarded. The remaining values are averaged to obtain HR sample. If all candidates are rejected, extrapolated values from last 3 s and 10 s are added to candidate set. The most extreme values (without comparison to previous sample) are discarded and the remaining set is averaged to produce HR. The final result is filtered using lowpass IIR filter. In order to compare HR signals from different devices we calculated the root mean squared error (RMSE) between the signals and a reference as well as a Pearson correlation coefficient between data gathered from all devices for one participant.

GSR processing GSR contains two components, the skin conductance level (SCL) (tonic level) and skin conductance response (SCR) known as phasic response. SCL is slowly changing signal dependent on individual factors such as hydration level or skin dryness. It is considered not to be informative on its own [?]. SCR is a set of alterations on top of SCL usually correlated with emotional responses to presented stimuli [?,?]. Galvanic skin response can be described in terms of skin resistance or conductance and can be expressed using different units. The first analysis step was to convert all the data to conductance in μS. The optimal GSR sampling rate is not clearly defined. Schmidt [?] recommends 15-20 Hz sampling rate and low pass filtering to 5 Hz, Ohme [?] proposes downsampling to 32 Hz without specifying the initial sampling rate, Nourbakhsh in his research used sampling rate of 10 Hz [?]. Sampling rates of MS Band 2 (5 Hz), Empatica E4 (4 Hz) and BITalino (1000 Hz) are fixed by the manufacturers so it was impossible adjust the values. E-Health acquired data using 32 samples per second. Finally, for data comparison we resampled all signals to 10 Hz.

To extract SCR, we estimated the tonic level and subtracted it from original signal. The SCL was obtained by extracting and interpolating local skin conductance minima. Finally, the SCL was smoothed using IIR lowpass filter [?]. To compare devices we calculated the Pearson correlation coefficient between recorded signals. Because each measurement was taken in the same time, on the same person the results should be highly correlated. This however does not provide an answer to the question which signal is better.

In order to assess which device produces better results we used following procedure. The SCR signal was deconvolved with single phasic response model [?,?] defined by equation $b[t] = e^{\frac{-t}{\tau_0}} - e^{\frac{-t}{\tau_1}}$. According to [?] recommended values for τ_0 and τ_1 are 2 and 0.75 respectively. The resulting signal is called driver function. It is smoothed using lowpass IIR filter. If SCR signal contains fragments similar to model response, corresponding high amplitude positive peaks appear in smoothed driver function, therefore if SCR signal contains data similar to model response, the driver function should be mostly positive. To examine it we calculate ratio of energy of positive values to whole signal (PTR):

$$PTR = 10 \log_{10} \left(\frac{RMS(max(0, DF))}{RMS(DF)} \right), \tag{1}$$

where DF is a driver function.

The SCR signal is reconstructed by convolving smoothed driver with model response. For noisy SCR signal smoothing the driver should alter the reconstructed signal significantly. To examine it we calculate signal-to-noise ratio:

$$SNR = 10 \log_{10} \left(\frac{SCR_r}{SCR_o - SCR_r} \right), \tag{2}$$

where SCR_r is the reconstructed signal and SCR_o is the original SCR data.

6 Results of Experiments

The experiment was conducted on 7 participants. We selected the fragments were data from all the devices was available, obtaining more than 2 hours of overlapping signals. Initial evaluation revealed that data gathered during a physical activity vary significantly more from the reference than the data from the initial part of the experiment. As our main interest is the quality of signals gathered while the participant is stationary, we decided to analyse both parts separately. The presented results refer to stationary part unless stated otherwise.

Correlation of HR signals is presented in Tab. 1. Each cell contains maximum, average and minimum value from the whole dataset. Average correlation between measured data and reference (Polar H6) is the best for MS Band 2. High spread for BITalino and Empatica E4 indicate that they are capable of producing better results, but they are very sensitive to device placement and measurement conditions. Moreover, the correlation factor for BITalino was decreasing with each experiment. Most likely it is caused by reusing the ECG electrodes, which should be replaced more frequently. The differences in correlation are mirrored in RMSE parameter (Tab. 2, lower values are better). Additionally, we observed that MS Band 2 and eHealth record HR with clearly visible quantization (Fig. 1).

Table 1. Heart rate signals correlation (${}_{max}/\text{average}/{}_{min}$)

	BITalino	eHealth	Empatica E4	MS Band 2	Polar H6
BITalino	${}_{1.00}/1.00/{}_{1.00}$	${}_{0.76}/0.13/{}_{-0.34}$	${}_{0.92}/0.48/{}_{-0.17}$	${}_{0.88}/0.17/{}_{-0.35}$	${}_{0.97}/0.18/{}_{-0.40}$
eHealth	${}_{0.76}/0.13/{}_{-0.34}$	${}_{1.00}/1.00/{}_{1.00}$	${}_{0.78}/0.28/{}_{-0.21}$	${}_{0.68}/0.42/{}_{0.19}$	${}_{0.84}/0.55/{}_{0.14}$
Empatica E4	${}_{0.92}/0.48/{}_{-0.17}$	${}_{0.78}/0.28/{}_{-0.21}$	${}_{1.00}/1.00/{}_{1.00}$	${}_{0.86}/0.32/{}_{-0.25}$	${}_{0.93}/0.41/{}_{-0.26}$
MS Band 2	${}_{0.88}/0.17/{}_{-0.35}$	${}_{0.68}/0.42/{}_{0.19}$	${}_{0.86}/0.32/{}_{-0.25}$	${}_{1.00}/1.00/{}_{1.00}$	${}_{0.88}/0.66/{}_{0.37}$
Polar H6	${}_{0.97}/0.18/{}_{-0.40}$	${}_{0.84}/0.55/{}_{0.14}$	${}_{0.93}/0.41/{}_{-0.26}$	${}_{0.88}/0.66/{}_{0.37}$	${}_{1.00}/1.00/{}_{1.00}$

Table 2. RMSE [dB] for heart rate signals (${}_{max}/\text{average}/{}_{min}$)

	BITalino	eHealth	Empatica E4	MS Band 2
RMSE [dB]	${}_{27.37}/13.67/{}_{1.70}$	${}_{8.53}/4.82/{}_{2.81}$	${}_{18.11}/6.97/{}_{1.53}$	${}_{7.81}/4.06/{}_{2.54}$

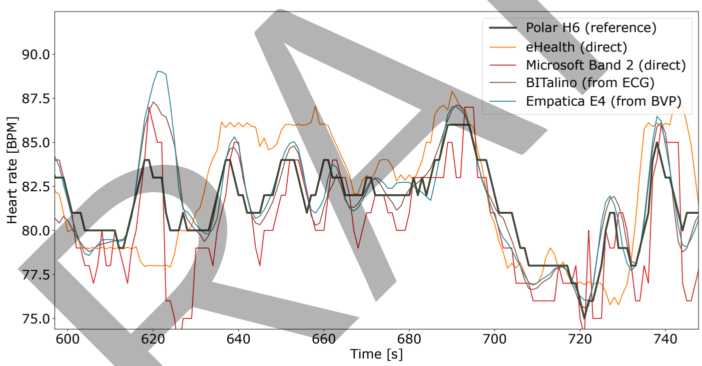

Fig. 1. Fragment of HR signals gathered using different devices

While working on the HR extraction algorithm we noticed that using shorter analysis window and skipping final smoothing reveals HR variability (HRV). This is a great advantage of BITalino and Empatica E4 over the remaining devices. According to [**?**], during inhale the HR is growing and while exhale it is decreasing. This information may be used to extract breathing frequency and respiratory sinus arrhythmia (RSA). In future experiments both parameters may be used to examine the emotional state of the subjects.

GSR signals correlation indicate high similarity of BITalino and eHealth (Tab. 3). The devices provide data with distinguishable phasic response and similar peak locations (Fig. 2). Signals from eHealth contain more noise than from BITalino, but they can be easily filtered due to high frequency and low amplitude of distortions. The amplitude of skin conductance response from Empatica E4 is lower than from eHealth and BITalino, but individual peaks are recognizable. Signals from Empatica have highest SNR and PTR (similarity to theoretical response) levels (Tab. 4). Unfortunately not all the peaks observed in results from BITalino and eHealth are present what is reflected in lower correlation factor. It may be caused by different sensor location (Empatica measures conductance GSR on wrist, eHealth and BITalino on fingers). GSR from MS Band 2 is uncorrelated with other devices and we were unable to extract any useful data. No phasic response can be observed therefore there is no point in analysing the remaining parameters.

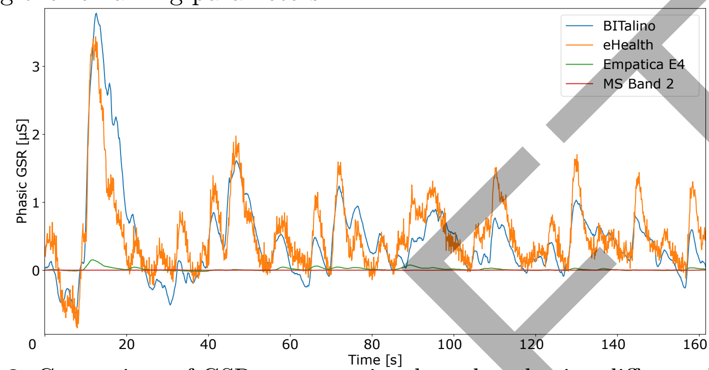

Fig. 2. Comparison of GSR response signals gathered using different devices

Table 3. Skin conductance response correlation ($_{max}$/average/$_{min}$)

	BITalino	eHealth	Empatica E4	MS Band 2
BITalino	1.00/1.00/1.00	0.84/0.70/0.41	0.55/0.30/0.05	0.19/0.06/-0.03
eHealth	0.84/0.70/0.41	1.00/1.00/1.00	0.64/0.32/0.02	0.06/0.04/0.02
Empatica E4	0.55/0.30/0.05	0.64/0.32/0.02	1.00/1.00/1.00	0.09/0.01/-0.09
MS Band 2	0.19/0.06/-0.03	0.06/0.04/0.02	0.09/0.01/-0.09	1.00/1.00/1.00

Table 4. Signal parameters for extracted skin conductance response ($_{max}$/average/$_{min}$)

	BITalino	eHealth	Empatica E4	MS Band 2
PTR [dB]	-0.19/-0.76/-2.80	-0.42/-0.73/-1.63	-0.14/-0.55/-1.68	0.00/-0.63/-1.58
SNR [dB]	23.13/18.07/13.32	16.51/12.59/4.84	26.80/19.99/3.52	20.48/9.23/0.22

The physical activity test was taken only by 4 out of 7 experiment participants. The results are compared to the stationary part of the experiment. Obtained parameters are expressed as a difference between the active and stationary part. In some cases the recorded HR was correlated with subject movement instead of the real HR. For BITalino and Emapatica, where HR was calculated by our own script it can probably be fixed by combining the ECG/BVP data with signals from acceleration sensor, but it was not tested yet. This problem affects results from MS Band 2 the most and eHealth the least (Tab. 5). Other issue occurs in GSR measurements, where during movement the contact between body and sensor was not constant what leads to sudden conductance changes. Both problems result in lower signal correlation for GSR and HR (Tab. 5, 6). The only case were the correlation has significantly grown is the GSR signal from MS Band 2 and Empatica E4, however it should be noted that the change was from correlation factor $-0,02$ to $0,20$, therefore the data is still weekly correlated.

Table 5. Difference between HR signals correlation during the activity test and the first part of the experiment ($_{max}$/average/$_{min}$)

	BITalino	eHealth	Empatica E4	MS Band 2	Polar H6
BI[5]	$_{0.00}$/0.00/$_{0.00}$	$_{-0.17}$/-0.14/$_{-0.21}$	$_{-0.58}$/-0.23/$_{0.14}$	$_{-0.48}$/-0.21/$_{-0.03}$	$_{-0.31}$/0.02/$_{0.41}$
eH	$_{-0.17}$/-0.14/$_{-0.21}$	$_{0.00}$/0.00/$_{0.00}$	$_{-0.38}$/-0.33/$_{0.21}$	$_{0.13}$/-0.14/$_{-0.23}$	$_{0.00}$/-0.31/$_{-0.35}$
E4	$_{-0.58}$/-0.23/$_{0.14}$	$_{-0.38}$/-0.33/$_{0.21}$	$_{0.00}$/0.00/$_{0.00}$	$_{-0.11}$/-0.17/$_{0.21}$	$_{0.03}$/-0.22/$_{0.14}$
B2	$_{-0.48}$/-0.21/$_{-0.03}$	$_{0.13}$/-0.14/$_{-0.23}$	$_{-0.11}$/-0.17/$_{0.21}$	$_{0.00}$/0.00/$_{0.00}$	$_{-0.41}$/-0.54/$_{-0.60}$
H6	$_{-0.31}$/0.02/$_{0.41}$	$_{0.00}$/-0.31/$_{-0.35}$	$_{0.03}$/-0.22/$_{0.14}$	$_{-0.41}$/-0.54/$_{-0.60}$	$_{0.00}$/0.00/$_{0.00}$

Table 6. Difference between skin conductance response correlation during the activity test and the first part of the experiment ($_{max}$/average/$_{min}$)

	BITalino	eHealth	Empatica E4	MS Band 2
BITalino	$_{0.00}$/0.00/$_{0.00}$	$_{-0.05}$/-0.20/$_{-0.32}$	$_{0.13}$/-0.02/$_{0.08}$	$_{0.13}$/0.00/$_{-0.14}$
eHealth	$_{-0.05}$/-0.20/$_{-0.32}$	$_{0.00}$/0.00/$_{0.00}$	$_{-0.03}$/-0.02/$_{0.16}$	$_{0.15}$/0.00/$_{-0.13}$
Empatica E4	$_{0.13}$/-0.02/$_{0.08}$	$_{-0.03}$/-0.02/$_{0.16}$	$_{0.00}$/0.00/$_{0.00}$	$_{0.34}$/0.22/$_{0.15}$
MS Band 2	$_{0.13}$/0.00/$_{-0.14}$	$_{0.15}$/0.00/$_{-0.13}$	$_{0.34}$/0.22/$_{0.15}$	$_{0.00}$/0.00/$_{0.00}$

Our critical analysis demonstrates, that Bitalino remains the most prospective platform for both HR and GSR measurements, especially that the technical support for the e-Health is being phased out. As for secondary HR readings MS Band can be used. While the Bitalino kit does not have the wristband form, it can be turned into a wearable using the Bitalino Freestyle kit and 3D-printed boxes. It is worth emphasizing, that the devices we selected in this paper offer real-time HR and GSR monitoring, as opposed to the vast majority of fitness trackers, e.g. from Fitbit, that offer only highly filtered and averaged data.

7 Conclusions and Future Work

This paper discusses the practical aspects of the construction of a measurement framework for affective computing and telemedicine based on low-cost, portable devices. We provide analysis of results of experiments aimed at critical comparison of quality of HR and GSR signals in selected devices.

In our future works on the framework we will consider focusing on the Bitalino, possibly combined with the MS Band 2 in some cases. Using the more reliable affective data acquired from these devices we will work on developing effective classification methods for the emotional condition of the user. These methods will be implemented on mobile devices, such as smartphones.

[5] BI – BITalino, eH – eHealth, E4 – Empatica E4, B2 – MS Band 2, H6 – Polar H6