Smart-Wearables and Heart-Rate Assessment Accuracy

André F. Manso[1], Ana L. N. Fred[2], Rui C. Neves[3] and Rui C. Ferreira[4]

Abstract— **Wearables and smart devices are making their way into people's life bringing constant connectivity. The inclusion of bio-sensors into these devices allows for a pervasive and constant monitoring of habits and also physical condition and physiological variables. In particular, seamless heart rate measurements can be of paramount importance in cardiac conditions, paving the way to personalized health care scenarios. However, reliability and accuracy of these sensor platforms remains an issue that is many times not properly addresses by manufactures. To shed some insight into this issue a smart device was placed against state-of-the-art equipment to assess its accuracy in heart rate measurements.**

I. INTRODUCTION

In recent years, wearables have been taking their place in biological signal acquisition. Sensors are becoming smaller, lighter, and cheaper as computing power is evermore affordable and small-sized. In particular, the inclusion of biosensors in smart devices has seen a spike, with purposes like biometrics or sports performance tracking. In this context, Photoplethysmography (PPG) is a major tool, being one of the less invasive techniques in heart rate (HR) estimation which makes it a very good candidate for inclusion into wearable devices [1]. Most Smartwatches nowadays include one of these sensors to allow user heart rate monitoring.

Continuous health monitoring is a very promising field of medical development [2] and this type of devices can play an important role in pervasive medical data collection. This allows physicians to monitor their patients in real time and collect data over long periods, paving the way to personalized health care scenarios. In particular, heart rate can be used as a major clinical indicator for patients with heart diseases [3], [4], [5].

Despite all the advantages and promises [6], wearables have some drawbacks that greatly affect the quality of their HR estimates. Smartwaches, in particular, position the PPG sensor in the distal portion of the posterior forearm. This is a location where PPG signal is present but faint due to reduced concentration of blood flow [7]. Another major problem is signal corruption by motion artifacts. This conundrum is present because these devices tend to be heavy enough to have their own dynamics i.e. they move independently from the forearm by inertia. In addition, it is not comfortable to have the device too tight to the skin, and excessive pressure reduces superficial blood flow, thus further reducing signal to noise ratio (SNR). In fact smartwaches and wearable sensors have been studied many times [6], [8] and even its applicability as a source of clinical information has been proposed [9], [10], [4]. However, accuracy of this type of devices has been questioned [9], [11], [12] and error margins for HR estimates produced by smartwaches were proven considerable.

When developing a platform for remote patient monitoring, for example, it is very relevant to have information about performance and technical aspects of sensors and algorithms included into commercial systems that manufacturers are most commonly not whiling to share. This makes it harder to choose the sensor platform, fitness tracker, smartwatch, etc... that fits better with project requirements, as reliable information on how a specific device behaves may be very important for platform selection.

Being Samsung one of the great players in the smartwaches and wearables field, a very recent smartwatch sensor platform, Samsung Gear S3 [13], was tested against an ECG-based ground truth device to provide a characterization of the error in HR determination.

Results are presented in the form of absolute error (AE) as it is an easy to read quantity providing a more informative measure, making easier for physicians to have a perception of measurements and the confidence interval provided by any novel system based on these devices.

II. METHODOLOGY

In order to evaluate the accuracy of HR estimates given by Samsung Gear S3, the later was used to record HR in different situations with several subjects with and without cardiac pathologies. Simultaneously a BITalino [14] and a medical-grade certified device were used to determine the HR from ECG data and provide a ground truth reference for Gear's values.

A. Data Acquisition

1) Samsung Gear S3 Frontier: A custom app was implemented with sensor interaction being made through the manufacturer provided API [15]. Data was sent to an android smartphone via Bluetooth for storage. The API allows to retrieve raw PPG data as well as HR calculated by the smartwatch from PPG data using manufacturer's algorithm. The

*This work was supported by the Portuguese Foundation for Science and Technology under grant PTDC/EEI-SII/7092/2014

[1]André F. Manso is with IST - Instituto Superior Técnico Universidade de Lisboa, Biomedical Engineering, Av. Rovisco Pais, 1, Lisboa, Portugal e-mail: andre.manso@ist.utl.pt

[2]Ana L. N. Fred is with Instituto de Telecomunicações, IST - Instituto Superior Técnico Universidade de Lisboa, Av. Rovisco Pais, 1, Torre Norte - Piso 10 1049-001 Lisboa, Portugal e-mail: afred@lx.it.pt

[3]Rui C. Neves is with CAST - Cons. e Apl. em Sistemas e Tecnologias, Lda., R. General Silva Freire 157 H, 1800-210 Lisbon, Portugal e-mail: rcneves@cast.pt

[4]Rui C. Ferreira is with the Cardiology Department of Hospital Santa Marta – CHLC, R. de Santa Marta 50, 1169-1024 Lisbon, Portugal e-mail: cruzferreira@netcabo.pt

device produces HR estimations at 25Hz, PPG is collected at 25Hz and tri-axial accelerometry is collected at 100Hz.

2) BITalino: BITalino is a customizable sensor platform that was used to collect one derivation of ECG at 1000Hz. This device was proven to be a reliable sensor with performances similar to medically certified devices [16], [14]. Based on BITalino, a chest band was designed and built using an elastic and adjustable band that sits under the breasts. This device in this location allows for a low-noise ECG acquisition even when the subject is moving or exercising.

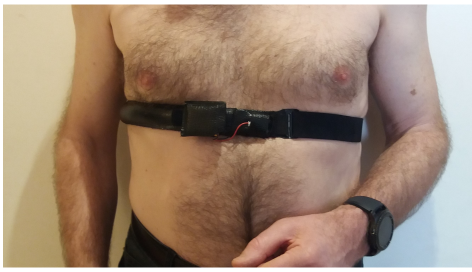


Fig. 1: Subject wearing chest band and smartwatch used for data acquisition.

3) Hospital Equipment: The device used is a Mortara XScribe 3.10.10 by Mortara Instrument [17] that collects 12 lead ECG, producing a very reliable HR estimate as it is a certified medical device.

B. Experimental Protocol

To validate heart rate values calculated by Gear S3 three experiments were conducted:

- E1 - Consisted on acquiring data for short periods of time with subjects performing a specific activity:
  - Resting state, moving as least as possible (2min)
  - Walking at a regular pace (5min)
  - one subject was asked to pedal on a stationary bike at moderate pace (5min)
  - one subject was asked to run at moderate pace (2min)

  Some of these acquisitions were repeated to verify reproducibility and infer on system robustness.
- E2 - On this experiment, data was recorded for periods between 1 and 2 hours while subjects performed their usual activities during their daily lifes.
- E3 - The experiment consisted on collecting data from 3 hospitalized patients with various cardiac diseases during a stress test, taking place at hospital. These tests consisted of patients walking on a treadmill, increasingly fast until they are not able to continue, followed by a rest period. Patients ECG and energy consumption is monitored while executing the task. The procedure was performed according to Bruce protocol [18].

C. Data Processing

After the experiments, data was processed in order to produce HR estimates from the acquired signals (PPG, ECG). To evaluate estimates quality, performance was measured having as ground truth HR values calculated from the ECG. For stress tests HR ground truth is provided by the hospital's equipment, while for the other experiments the ground truth is calculated from data collected by the BITalino chest band.

ECG data processing consisted on the following. First, the R peaks were detected using method proposed in [19], from which the instantaneous Heart Rate was computed using eq. (1). Since the smartwatch does not provide instantaneous HR estimates, but rather provides HR values in a 25Hz frequency, the ECG-based HR for consecutive, non-overlaping, 10s windows was determined as the median of the instantaneous HR values for each window. The HR provided by Gear S3 was also determined as the median of consecutive, non-overlaping, 10s windows from the 25Hz samples it returns.

$$HR_i = \frac{60 * f_s}{t_{R_{i+1}} - t_{R_i}} \tag{1}$$

HR_i → ith instantaneous HR
f_s → Sensor's sampling frequency
t_{R_i} → sample number of the ith R-peak

D. Accuracy Measurement

The accuracy of the HR values produced by the smartwatch, and also the ones estimated from the ECG signal, was summarized as statistics of the absolute error (AE) e.g. mean absolute error (MAE) as defined in eq. (2).

$$MAE = \frac{1}{N} \sum_{i=0}^{N} \left| HR_i^{true} - HR_i^{est} \right| \tag{2}$$

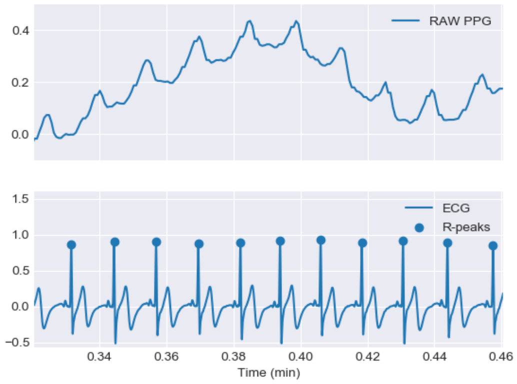
N → nr. of HR values estimated for a subject
HR_i^{true} → Ground-truth value of instantaneous HR
HR_i^{est} → Estimated value of instantaneous HR

III. Experimental Setup and Results

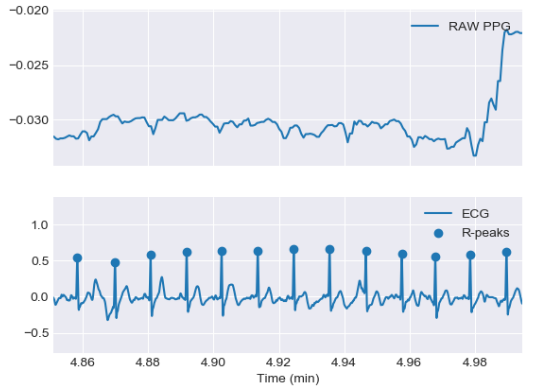
Data was collected from two separate groups performing different activities. Stress tests were undertaken by 3 patients with various cardiac conditions and ages 42, 47 and 50 years old. The other experiments enumerated in section II-B were performed by 3 volunteers all 23 year olds. In following figures and tables Si denotes subject i and Tj denotes jth trial.

After data collection from subjects, the first observation of the PPG signal revealed the vast influence of motion artifacts (MA). This is depicted in fig. 2. When the subject is completely immobile it is easy to find, by inspection, a correlation between ECG and PPG signal peaks (fig. 2a) corresponding to systolic peaks. The same is not true when the subject is moving, as signal corruption greatly affects PPG signal and correlation is no longer obvious and maybe not present at all (fig. 2b).

Observing fig. 5, it is visible that Gear S3 is not capable of keeping up with fast variations in HR and fig. 7 clearly illustrates the difficulty in determining HR accurately during movement, with a very large estimation error being present, until subject enters the resting phase of the stress test.



(a) Raw signal while completely immobile. HR calculated from PPG (first curve): 75bpm; HR calculated from ECG (second curve): 79bpm.



(b) Raw signal while walking. HR calculated from PPG (first curve): 52bpm; HR calculated from ECG (second curve): 81bpm

Fig. 2: Segments of raw signal captured by the sensors in different conditions with the synchronous identified R-peaks.

A. *Short Acquisitions (E1)*

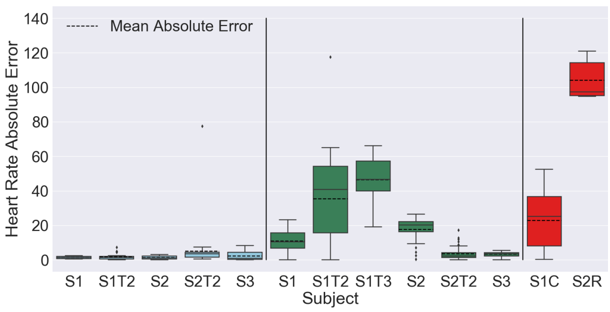


Fig. 3: Absolute Error of HR while performing different activities: (left) resting; (middle) walking; right cycling and running.

When analyzing the absolute error of HR determined from data collected while subjects were performing specific activities it is very obvious that motion highly corrupts sensor data and thus greatly damages accuracy. In fig. 3 is very clear a tendency to error increase as subjects go from resting to walking, cycling or running. Another thing that can be noted in fig. 3 and table I is the relatively large difference in the error values between subjects and trials. This may indicate low robustness of the system, as it is affected by sensor positioning and tightness and also by how the subjects move, as some individuals present more arm movement which can further disrupt accuracy.

TABLE I: Mean Absolute Error (MAE) of each experiment and activity averages.

Activity	Subject	MAE	Average
Resting	S1	1.4	2.4
	S1T2	1.8	
	S2	1.3	
	S2T2	5.1	
	S3	2.2	
Walking	S1	10.8	21.2
	S1T2	35.3	
	S1T3	46.4	
	S2	17.7	
	S2T2	3.4	
	S3	3.3	
Cycling	S1	22.8	63.4
Running	S2	104.1	

B. *Daily life (E2)*

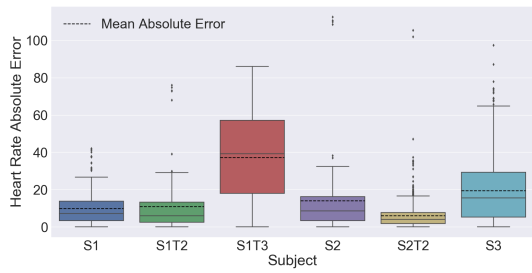


Fig. 4: Absolute Error of HR while subjects perform their usual life activities (office working, walking, eating, etc...).

TABLE II: Mean Absolute Error (MAE) of each subject during daily life activities.

Activity	Subject	MAE	Average
Daily-life activities	S1	9.7	16.2
	S1T2	10.8	
	S1T3	37.1	
	S2	13.9	
	S2T2	6	
	S3	19.3	

Table II and fig. 4 demonstrate clearly that during daily-life activities like walking or any activity that implies arm movement, Gear S3 has a poor performance as a sensor platform and in certain occasions it is not even possible to detect any type of signal trends. In a medical context, this could lead to gross errors when trying to diagnose or follow a patient, or it would require further tests which renders the pervasive monitoring less efficient or even counterproductive.

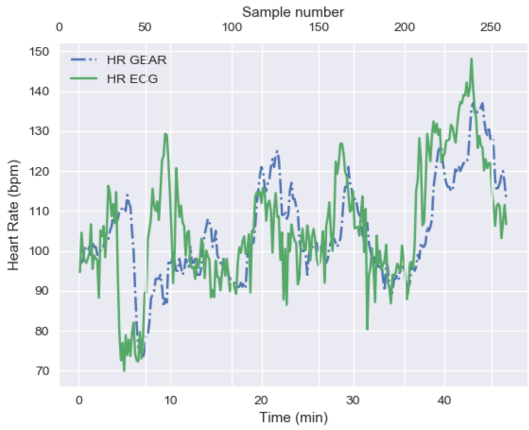


Fig. 5: Example of HR curves obtained during daily life activities.

C. Stress Test (E3)

As expected, results during a stress test, where the subject is moving with some intensity, HR determination error is very large and as can be seen in fig. 7, error is specially large during exercise part of the stress test, supporting hypothesis of motion artifacts corrupting the signal. Error during these tests reached >100bpm wich renders this sensor platform completely unfit for this kind of context.

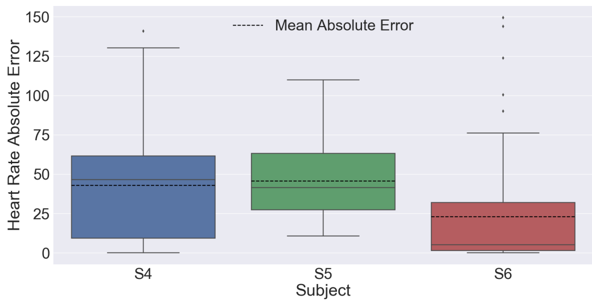


Fig. 6: Absolute Error of HR during stress test.

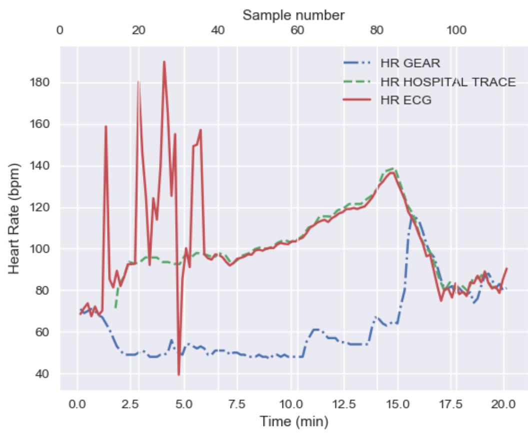


Fig. 7: Example of HR curve obtained during stress test.

IV. Discussion

Analyzing the results obtained, the first and main aspect to point out is that the PPG signal collected with Samsung Gear S3 is very easily corrupted by motion artifacts. Even with mild movement, the signal is not reliable and systolic peaks can no longer be identified. This renders the device unfit for continuous monitoring, since HR estimation from PPG data has very low accuracy.

Probable causes for this poor performance are the fact that only one source of light and one detector are present, the weight of the device and the development cost.

Samsung's Gear S3 sensor has a single LED with a single wavelength. Some other PPG sensors use two LEDs and some two wavelengths which eases the signal processing and improves performance. The existence of a single source of light, with only one wave length, implies that less information is gathered and the signal processing options are somehow limited.

Concerning the weight, it must be taken into account that this is a very complex device that can perform endless tasks with native support for 3rd party apps. All these functionalities and capabilities make the device heavier and lager, thus, making it harder to keep comfortably and tightly secure to the wrist without it having its own dynamics. This is a problem as the relative movement between the sensor and the subject is the sole cause for movement artifacts in the signal. In addition the positioning of the sensor in an area with a relatively low SNR greatly increases the impact of motion artifacts. Also this device has the sensor at the same level of the watch back plate, whereas other devices place the sensor in an elevation of the back plate, which makes the sensor to be slightly pressed against the skin and increases SNR, and thus, accuracy.

Due to poor quality PPG signal acquired, several algorithms were used in an attempt to get better HR estimates. Algorithms used included adaptive filtering with and without Laguerre expansion [20], [21], signal separation by sparse signal reconstruction [22], [23] and onset detection [24]. A total of 8 algorithms were used to process the PPG and accelerometry data coming from S3 to produce estimates of HR. However, the results obtained using all this algorithms performed equally bad, or even worse, than Gear's algorithm and for this reason, they were not mentioned previously. This clearly indicates that elevated error in Gear's HR estimations is probably related with a low quality signal and not with a poorly performing algorithm.

V. Conclusions

In this work the accuracy of HR estimates produced by a wearable smartwach was studied using by Samsung Gear S3. Analyzing the results obtained, it is fair to conclude that this system, and this type of commercial wearables, are not yet fit for long term monitoring to be used by physicians in either diagnostics nor patient follow up. This type of application would require a precision larger than what is possible to achieve with current devices and algorithms.

As researchers and manufacturers continue to improve device and algorithms, it is expectable that in a near future patients and physicians benefit from more accurate devices to provide better health-care.

REFERENCES

[1] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, no. 3, p. R1, 2007.

[2] C. S. Bloss, N. E. Wineinger, M. Peters, D. L. Boeldt, L. Ariniello, J. Y. Kim, J. Sheard, R. Komatireddy, P. Barrett, and E. J. Topol, "A prospective randomized trial examining health care utilization in individuals using multiple smartphone-enabled biosensors," *PeerJ*, vol. 4, p. e1554, 2016.

[3] K. H. Shelley, "Photoplethysmography: beyond the calculation of arterial oxygen saturation and heart rate," *Anesthesia & Analgesia*, vol. 105, no. 6, pp. S31–S36, 2007.

[4] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, no. 3, p. R1, 2007.

[5] H. B. Abramowitz, L. A. Queral, W. R. Flinn, P. F. Nora, L. K. Peterson, J. J. Bergan, and J. S. Yao, "The use of photoplethysmography in the assessment of venous insufficiency: a comparison to venous pressure measurements," *Surgery*, vol. 86, no. 3, pp. 434–441, 1979.

[6] R. Rawassizadeh, B. A. Price, and M. Petre, "Wearables: Has the age of smartwatches finally arrived?" *Communications of the ACM*, vol. 58, no. 1, pp. 45–47, 2015.

[7] S. Lee, H. Shin, and C. Hahm, "Effective ppg sensor placement for reflected red and green light, and infrared wristband-type photoplethysmography," in *Advanced Communication Technology (ICACT), 2016 18th International Conference on*. IEEE, 2016, pp. 556–558.

[8] G. Bieber, M. Haescher, and M. Vahl, "Sensor requirements for activity recognition on smart watches," in *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2013, p. 67.

[9] F. El-Amrawy and M. I. Nounou, "Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial?" *Healthcare informatics research*, vol. 21, no. 4, pp. 315–320, 2015.

[10] A. Carpenter and A. Frontera, "Smart-watches: a potential challenger to the implantable loop recorder?" *Europace*, vol. 18, no. 6, pp. 791–793, 2016.

[11] J. Parak and I. Korhonen, "Evaluation of wearable consumer heart rate monitors based on photopletysmography," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2014, pp. 3670–3673.

[12] E. E. Dooley *et al.*, "Measuring the validity of self-monitoring heart rate and activity tracking wearables," Ph.D. dissertation, 2016.

[13] Samsung, "Samsung Gear S3." [Online]. Available: http://www.samsung.com/pt/wearables/gear-s3/

[14] J. Guerreiro, R. Martins, H. Silva, A. Lourenço, and A. L. Fred, "Bitalino-a multimodal platform for physiological computing." in *ICINCO (1)*, 2013, pp. 500–506.

[15] T. Project, "Tizen Sensor API," 2012. [Online]. Available: https://developer.tizen.org/development/api-references/native-application?redirect=/dev-guide/2.3.1/org.tizen.native.mobile.apireference/group__CAPI__SYSTEM__SENSOR__MODULE.html

[16] D. Batista, H. Silva, and A. Fred, "Experimental characterization and analysis of the bitalino platforms against a reference device," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*. IEEE, 2017, pp. 2418–2421.

[17] M. Instruments, "Mortara XScribe." [Online]. Available: https://www.mortara.com/products/healthcare/hospitals/cardiac-stress-testing/xscribe/

[18] M. W. Luong, M. Ignaszewski, and C. Taylor, "Stress testing: A contribution from dr robert a. bruce, father of exercise cardiology," *British Columbia Medical Journal*, vol. 58, no. 2, pp. 70–76, 2016.

[19] A. Lourenço, H. Silva, P. Leite, R. Lourenço, and A. L. Fred, "Real time electrocardiogram segmentation for finger based ecg biometrics." in *Biosignals*, 2012, pp. 49–54.

[20] P. T. Gibbs, L. B. Wood, and H. H. Asada, "Active motion artifact cancellation for wearable health monitoring sensors using collocated mems accelerometers," in *Proc. SPIE*, vol. 5765, 2005, pp. 811–819.

[21] L. B. Wood and H. H. Asada, "Active motion artifact reduction for wearable sensors using laguerre expansion and signal separation," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. IEEE, 2006, pp. 3571–3574.

[22] Z. Zhang, "Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction," *IEEE transactions on biomedical engineering*, vol. 62, no. 8, pp. 1902–1910, 2015.

[23] Z. Zhang, Z. Pi, and B. Liu, "Troika: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 522–531, 2015.

[24] R. Blazek and C. Lee, "Multi-resolution linear model comparison for detection of dicrotic notch and peak in blood volume pulse signals," *Anal Biomed Signals Images*, vol. 20, pp. 378–386, 2010.