The 10th IEEE International Conference on E-Health and Bioengineering - EHB 2022
Grigore T. Popa University of Medicine and Pharmacy, Iasi, Romania, November 17-18, 2022

Smartphone PPG Validation for a Depression Assessment Protocol

Diego Cajal[1,2], Mar Posadas-de Miguel[3], Concepción de la Cámara[3,4,5],
Spyridon Kontaxis[1,2], Jesús Lázaro[1,2], Raquel Bailón[1,2]

Affiliation 1: Biomedical Signal Interpretation and Computational Simulation (BSICoS) Group, Aragón Institute of Engineering Research (I3A), IIS Aragón, University of Zaragoza, 50018 Zaragoza, Spain
Affiliation 2: Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), 28029 Madrid, Spain
Affiliation 3: IIS Aragón Zaragoza
Affiliation 4: CIBER de Salud Mental.
Affiliation 5: Hospital Clínico Universitario Lozano Blesa

***Abstract*—A new smartphone application for pulse photoplethysmography (PPG) recording is validated for a depression assessment protocol, consisting of four stages that includes recordings of movement-related tasks, talking-related tasks, and stillness. The separation of these stages allows to evaluate the agreement between the PPG acquired using the smartphone camera and a commercial pulse oximeter, used as reference, in different conditions. Several time-domain and frequency-domain heart rate variability (HRV) metrics have been studied. Results suggest a strong agreement during stillness phases ($r \geq 0.96$ ($p < 0.05$) for all the HRV metrics), as well as a moderate downgrade during movement and talking. The agreement is also different between low-frequency related metrics ($r \geq 0.92$ ($p < 0.05$)) and high-frequency related metrics ($r < 0.78$ ($p < 0.05$)). The overestimation of the high-frequency component of the HRV may be a limitation for smartphone PPG in depression monitoring.**

Keywords—smartphone; PPG; depression; stress; validation.

I. Introduction

Pulse photoplethysmography (PPG) signals acquired from smartphone cameras have been investigated in a variety of applications, such as glucose level classification [1], vascular aging prediction [2], biometric authentication [3], and stress assessment [4]. Its usefulness in the fields of medicine, wellness, and sports, together with the fact that it can be obtained in a non-invasive manner, has made this signal an important object of study. Furthermore, PPG signal can be recorded by the built-in hardware of smartphones (flashlight and camera), and the ubiquity of these devices makes them very interesting in some applications. However, the quality of smartphone-camera-acquired PPG (SCPPG) signals is lower than the quality of signals obtained with conventional pulse photoplethysmographs. Movements, together with changes in finger pressure on the camera, are the main causes of artifacts. Therefore, their use for different applications must be validated within each scenario.

In this work, the validity of SCPPG is investigated during a depression assessment protocol that aims to induce stress in the subjects. The physiological response to stress has been shown to be different in depressed patients from in healthy subjects, which can be monitored by pulse rate variability (PRV) [5] and morphology-derived metrics [6].

II. Dataset and Methods

A. Dataset

10 healthy Caucasian subjects (age mean±std: 39±11; 4 women) underwent a depression assessment protocol (see Table I) consisting of: i) 5 minutes of basal state; ii) a Trail Making Test (TMT); iii) a Stroop Test (ST); iv) and 5 minutes of recovery. Both the TMT and the ST are stress-inducing tasks and were performed in a tablet.

During the basal stage, subjects were instructed by a guided relaxation audio. Afterwards, they performed a TMT, consisting of a first page of random positioned numbers that they must follow in ascending order using the index finger of their dominant hand and without raising the hand from the tablet. A second page is formed by numbers and letters, which must be followed alternating numbers and letters (1-A-2-B-3-C, and so on). The ST consisted of three pages. The first is formed by the words “red”, “green” and “blue” in a random order and black ink that subjects had to read. The second page is formed by colors (red, green, and blue) that subjects had to name. The third is formed by the same words, written with an ink that do not correspond to the word. Subjects had to name the color of the ink instead of the word written. Finally, a recovery stage consisted in a not guided relaxation right after the ST.

978-1-6654-8557-9/22/$31.00 ©2022 IEEE

TABLE I. STAGES OF THE PROTOCOL

Stage	*Duration*	*Type of activity*
Basal	5 minutes	Stillness
TMT	Until end	Movement
ST	Until end	Talking
Recovery	5 minutes	Stillness

The subjects were seated and instructed to limit their movements during the entire protocol, and to hold the smartphone (Xiaomi Pocophone F1, China) with their non-dominant hand while they covered the camera with the index finger. The camera layout in this device allows holding the smartphone in a comfortable position while not directly touching the flashlight, which can produce unpleasant heat in long recordings. Instead, the flashlight is at the right side of the camera at approximately 5mm distance (Fig. 1).

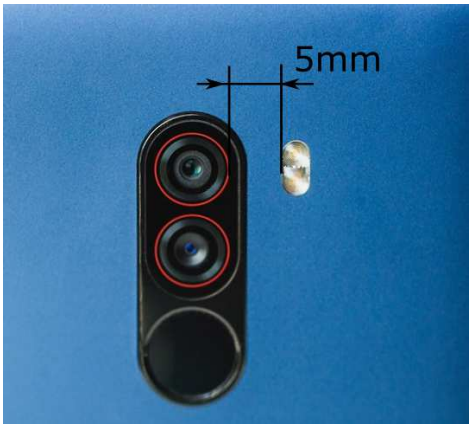

Fig. 1. Camera layout of the Xiaomi Pocophone F1

The flashlight was on during the recordings, and both the autofocus and autoexposure functions locked. This step is critical because these functions would cause non-physiological oscillations in the acquired SCPPG signal. These oscillations usually have a frequency similar to that of the pulses and can lead to confusing them. A Medicom system (Medicom MTD, Russia) was used simultaneously to record a conventional PPG signal in the ring finger of the same hand, for reference purposes.

The smartphone application used was self-developed for PPG recording by using Flutter (Google LCC., USA). The application processes the camera feed –with a resolution of 320x240 pixels, 24 frames per second, and RGB codification– by summing the values of the green component of each frame, thus obtaining a signal proportional to the intensity of the image. This signal is approximately –frame rate is dependent of the load of the operative system– sampled at 24 Hz. This signal is then upsampled using cubic splines at 250 Hz, matching the reference sample rate. Interpolation and further processing were performed offline by using MATLAB (MathWorks, USA).

B. Pulse detection and processing

Pulses were detected using the adaptative threshold algorithm described in [7] for each subject and protocol stage. This algorithm obtains the event series (t_k), *i.e.*, the timestamps of pulse occurrences.

SCPPG signals are frequently affected by motion artifacts, producing both false positives and false negatives in the event series. Thus, before any PRV further analysis, these errors must be detected and corrected. First, pulse-to-pulse interval series were computed using the interval function $d_{IF}(t_k)$, defined as

$$d_{IF}(t_k) = \sum_{k=1}^{K} (t_k - t_{k-1})\delta(t - t_k) \qquad (1)$$

Each event occurring at time t_k is represented by a unit impulse function $\delta(t-t_k)$ scaled by the length of the preceding interval. False positives produce an abrupt shortening of this scaling due to the introduction of an additional pulse between two actual pulses. A moving median of 30 samples is used to detect these outliers. The moving median produces an expected pulse-to-pulse interval (EPPI) at each t_k:

$$EPPI(t_k) = median(\{d_{IF}(t_i) | i \in \mathbb{N}, (k-15) < i \leq (k+15)\}) \qquad (2)$$

The interval at t_k is considered as a false positive if $d_{IF}(t_k) < (0.7xEPPI(tk))$. These false positives are deleted from the t_k series and $d_{IF}(t_k)$ is computed again. On the other hand, the interval at t_k is considered as a false negative if $d_{IF}(t_k) > (1.3xEPPI(tk))$. The treatment of false negatives was not the same for all the studied PRV metrics based on the results of [8], and it is detailed in Section II-C.

C. PRV metrics

For each protocol stage and subject, five PRV metrics are computed: mean heart rate (MHR), standard deviation of normal-to-normal interval (SDNN), root mean square of successive differences (RMSSD), low frequency power (P_{LF}) and high frequency power (P_{HF}).

The detected gaps in the t_k series were filled by an algorithm described in [8] before P_{LF} and P_{HF} computation. Then, the instantaneous heart rate is computed at 4 Hz using the integral pulse frequency modulation (IPFM) model [9]. Power spectral density of the instantaneous heart rate is estimated using periodograms after detrending with a 4th-order Butterworth high-pass filter with cutoff frequency 0.04 Hz. Periodograms are estimated from 0 to 0.4 Hz using 2^{10} points. Then, low-frequency and high-frequency powers are computed using trapezoidal numerical integration within the classic bands: 0.04 to 0.15 for the low-frequency component and 0.15 to 0.4 for the high-frequency component.

According to the results obtained in [8], gap filling after outlier removal is not convenient in the case of the time-domain metrics (MHR, SDNN and RMSSD). Thus, these indices were computed from $d_{IF}(t_k)$ after removal of outliers in the t_k series, with no additional gap filling.

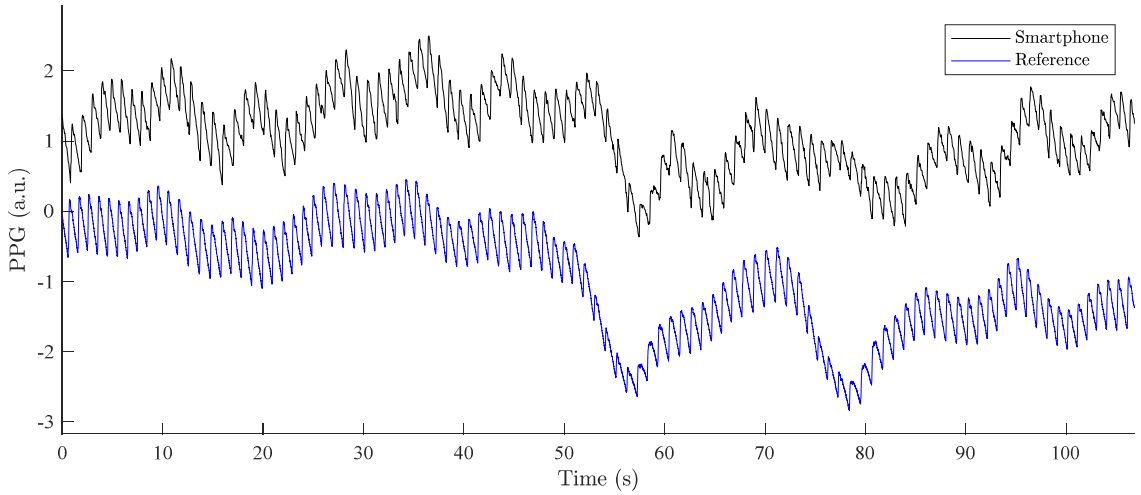

Fig. 2. PPG signals. SCPPG in black, Reference (Medicom) in blue.

D. Statistical analysis

PRV metrics obtained from smartphone PPG were compared to the Medicom system reference. Relative error ($\boldsymbol{\varepsilon}_r$) is calculated as the absolute value of metric differences, normalized by the reference values:

$$\epsilon_r = \left|\frac{smartphoneValues - referenceValues}{referenceValues}\right| \qquad (3)$$

The median and median absolute deviation (MAD) of $\boldsymbol{\varepsilon}_r$ is shown, both expressed as a percentage. Correlation is tested with the Pearson correlation coefficient (r) with significance level $\alpha = 0.05$. As it is strongly affected by outliers, cases with relative error greater than three times the median relative error have been removed before computing coefficients. In addition, the Bland-Alman's mean difference (Δ) is shown, expressed in percentage. This value is useful to assess if differences are biased, as well as the sign of the bias.

III. RESULTS

Table II shows a summary of the relative errors, correlation coefficients, and mean difference of each metric and protocol stage. An example of SCPPG it is shown in Fig. 2. Pearson correlation obtained $p < 0.05$ for all the cases. All metrics obtained the best results in the basal stage, followed by recovery, ST, and TMT, both in terms of relative error and correlation. All the metrics obtained $r \geq 0.96$ ($p < 0.05$) in the basal stage.

MHR showed almost perfect agreement, with $\boldsymbol{\varepsilon}_r \leq 0.28\%$ and $r = 1$ ($p < 0.05$) for all stages. Both SDNN and P_{LF} obtained errors lower than 9.32% and correlation coefficients over than 0.88. This contrasts with RMSSD and P_{HF} results. RMSSD error ranges from 32.54% to 54.40%, while P_{HF} error ranges from 28.47% to 54.83%. In addition, correlation coefficients are lower in comparison with other metrics, being $r < 0.76$ ($p < 0.05$) for RMSSD during the stress-inducing tasks, *i.e.*, TMT and ST.

TABLE II. RELATIVE ERROR, PEARSON CORRELATION COEFFICIENT AND MEAN DIFFERENCE OF EACH METRIC AND PROTOCOL STAGE

MHR	***Median ε_r (%)***	***MAD ε_r (%)***	***r***	***# Outliers***	***Δ (%)***
Basal	0.04	0.03	1.00	2	0.04
TMT	0.28	0.15	1.00	1	0.17
ST	0.08	0.05	1.00	2	0.03
Recovery	0.08	0.04	1.00	3	0.00
SDNN					
Basal	4.86	2.66	1.00	1	5.3
TMT	9.32	5.38	0.88	2	4.4
ST	6.78	3.72	0.99	1	6.2
Recovery	4.36	2.70	1.00	3	4.0
RMSSD					
Basal	39.17	30.15	0.96	3	25
TMT	36.73	26.77	0.68	0	38
ST	54.40	25.91	0.76	1	40
Recovery	32.54	23.58	0.96	3	25
P_{LF}					
Basal	1.68	1.29	1.00	2	0.21
TMT	4.97	3.46	1.00	4	1.2
ST	4.00	2.30	1.00	2	-0.1
Recovery	3.32	1.66	1.00	1	0.83
P_{HF}					
Basal	43.81	39.40	0.99	1	36
TMT	28.47	23.84	0.91	2	12
ST	44.18	27.91	0.98	3	25
Recovery	54.83	37.04	1.00	3	30

Results showed a clear difference between the basal/recovery and ST/TMT stages of the protocol, both in terms of correlation and relative error. The worst case is RMSSD during the TMT, where $r = 0.68$ ($p < 0.05$).

The number of outliers ranged from 1 to 4. The major number of outliers is found when computing frequency-domain metrics. Mean differences were positive in all the cases except for P_{LF} during the ST. MAD of $\boldsymbol{\varepsilon}_r$ followed the same trends as the median $\boldsymbol{\varepsilon}_r$.

IV. Discussion

The application has demonstrated almost a perfect correlation with the reference during the basal stage. Subjects were instructed to restrict their movements during this phase, while during the stress-inducing tasks, ST, and TMT, they need to talk or move, respectively. Thus, it was expected a downgrade in the stress-inducing task agreement. However, this downgrade is negligible in the MHR case ($\boldsymbol{\varepsilon}_r \leq 0.28\%$, $r = 1$), and it is low in SDNN ($\boldsymbol{\varepsilon}_r \leq 9.32\%$, $r \geq 0.88$) and P_{LF} ($\boldsymbol{\varepsilon}_r \leq 4.97\%$, $r = 1$). There is only a substantial downgrade both in RMSSD ($\varepsilon \leq 54.40\%$, $r \geq 0.68$) and P_{HF} ($\boldsymbol{\varepsilon}_r \leq 54.83\%$, $r \geq 0.91$). Recovery stage is the second best in terms $\boldsymbol{\varepsilon}_r$ and correlation. There are only slight differences with respect to the basal stage. Only the number of outliers increased significantly, especially for P_{HF}. It is interesting to note also that correlation is lower during the TMT (movement-related task) than during the ST (talking-related task), as it suggests that movements of the dominant hand may produce non-negligible movements in the non-dominant hand.

One important finding of this research is the fact that errors in low-frequency-related metrics (SDNN and P_{LF}) are significantly smaller than those of the high-frequency-related metrics (RMSSD and P_{HF}). The power of the high-frequency component of the PRV is overestimated by the smartphone PPG, as it can be seen in the mean difference of P_{HF} ($\Delta \geq 26$) and RMSSD ($\Delta \geq 30$).

Several subjects showed difficulties in standing still during the whole protocol. Movements ranged from small comfort adjustments or isolated hand spasms to continuous trembling. The last is, without any doubt, the most insidious case in this study, producing most of the outliers. While comfort adjustments and spams last for few seconds, trembling may affect the whole recording. Trembling can be absorbed by pulse oximeters, but it is transferred to the signal completely in the case of smartphone PPG, producing signals with apparently good quality. Pulses can be detected in these signals, but there are small shifts due to the trembling distortion. It seems reasonable that this trembling occurs within the high-frequency range of the PRV, thus affecting this component the most. On the other hand, small motion artifacts can produce bursts of missing data that can be corrected [8].

Overestimation of the high-frequency component is one of the main potential limitations of the smartphone PPG in depression assessment, as is one of the features with a best-demonstrated relationship with this disorder [10]. However, the strong correlation found suggests that these metrics may be still useful, as HF-related metrics may still discriminate stress if their overestimation is consistent in the different stages. Further studies will be necessary to: i) verify the overestimation of HF in more subjects; and ii) test whether this overestimation influences depression assessment.

V. Conclusion

A smartphone application for PPG recording has been validated during a depression assessment protocol. Agreement with the reference is almost perfect when the subject is stood still and relaxed. Errors arise when the subject is performing stress-inducing tasks, especially during movement-related tasks, although agreement still high. Smartphone PPG overestimates the power of the high frequency component of the PRV, being a potential limitation for depression assessment.

Acknowledgment

The computation was performed at the High Performance computing platform of NANBIOSIS ICTS. This work was supported by Ministerio de Ciencia e Innovación (PID2021-126734OB-C21), by Aragon Government and FEDER through BSICoS group (T39 20R), and by CIBER in Bioengineering, Biomaterials and Nanomedicine through Instituto de Salud Carlos III.

References

[1] Y. Zhang, Y. Zhang, S.A. Siddiqui, A. Kos. “Non-invasive blood-glucose estimation using smartphone PPG signals and subspace kNN classifier”. Elektrotehniski Vestnik, 86(1/2), 2019, pp. 68-74.

[2] L. Dall’Olio, N. Curti, D. Remondini, Y. Safi Harb. F.W. Asselbergs, G. Castellani, H.W. Uh. “Prediction of vascular aging based on smartphone acquired PPG signals”. Scientific reports, 10(1), 2020, pp. 1-10.

[3] G. Lovisotto, H. Turner, S. Eberz, I. Martinovic. “Seeing red: PPG biometrics using smartphone cameras”. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 818-819.

[4] D. Hernando, S. Roca, J. Sancho, Á. Alesanco, R. Bailón. “Validation of the apple watch for heart rate variability measurements during relax and mental stress in healthy subjects”. Sensors, 18(8), 2018, p. 2619.

[5] R. Castaldo, P. Melillo, U. Bracale, M. Caserta, M. Triassi, L. Pecchia. “Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis”. Biomedical Signal Processing and Control, 18, 2015, pp. 370-377.

[6] S. Kontaxis. E. Gil, V. Marozas, J. Lázaro, E. Garcia, M. Posadas-de Miguel, R. Bailón. “Photoplethysmographic waveform analysis for autonomic reactivity assessment in depression”. IEEE Transactions on Biomedical Engineering, 68(4), 2020, pp. 1273-1281.

[7] J. Lázaro, E. Gil, J.M. Vergara, P. Laguna. “Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children”. IEEE journal of biomedical and health informatics, 18(1), 2013, pp. 240-246.

[8] D. Cajal, D. Hernando, J. Lázaro, P. Laguna, E. Gil, R. Bailón. “Effects of missing data on heart rate variability metrics. Sensors, 22(15), 2022, p. 5774.

[9] J. Mateo, P. Laguna. “Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal”. IEEE Trans. Biomed. Eng., 50, 2003, pp. 334–343.

[10] L.M. Holsen, J.H. Lee, S.B. Spaeth, L.A. Ogden, A. Klibanski, S. Whitfield-Gabrieli, J.M. Goldstein. “Brain hypoactivation, autonomic nervous system dysregulation, and gonadal hormones in depression: a preliminary study. Neuroscience letters”, 514(1), 2012, pp.57-61.